

From: Ramirez, Delia  
Sent: Tuesday, November 20, 2001 9:48 AM  
To: STIC-Biotech/ChemLib  
Subject: 09/854,844 search

MEJ

Hi,

I would like to request the following searches (09/854,844 Hu et al.):

1. a standard search of seq id 1 in the nucleic acid databases (commercial and interference)
2. a standard search of seq id 2 in the protein databases (commercial and interference)
3. an oligo search of seq id 1 in the nucleic acid databases (commercial and interference)

Thank you,

Delia M. Ramirez, Ph.D.  
Patent Examiner - Art Unit 1652  
USPTO  
1911 S. Clark Street, Crystal Mall 1, 10D04, Mail room 10C01  
Arlington, VA 22202  
(703) 306-0288  
delia.ramirez@uspto.gov

RECEIVED  
NOV 20 2001  
STIC

nucl-1  
prot-2

Searcher: P. Schurber  
Phone: 308-4292  
Location: CM 12C14  
Date Picked Up: \_\_\_\_\_  
Date Completed: 11/26  
Searcher Prep/Review: 5  
Clerical: \_\_\_\_\_  
Online time: 7

TYPE OF SEARCH:  
NA Sequences: 2  
AA Sequences: 1  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST(where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: Compage  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

BEST AVAILABLE COPY

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2001, 01:54:41 ; Search time 30.98 Seconds  
(without alignments)  
251.328 Million cell updates/sec

Title: US-09-854-844-2  
Perfect score: 1863  
Sequence: 1 MGPAGCAFTLLLLIGISVCG.....GRELTGPELLTGLDGIYNLK 346

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2.6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2.6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2.6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2.6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2.6/ptodata/2/1aa/FCUS\_COMB.pep:\*  
6: /cgn2.6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	526	28.2	314	4 US-09-008-271A-3	Sequence 3, Appl
2	504	27.1	299	4 US-08-944-483-66	Sequence 66, Appl
3	480	25.8	276	2 US-09-016-366A-15	Sequence 15, Appl
4	480	25.8	276	2 US-08-978-404B-21	Sequence 21, Appl
5	479.5	25.7	270	2 US-08-978-404B-8	Sequence 8, Appl
6	468	25.1	274	2 US-08-978-404B-5	Sequence 5, Appl
7	467	25.1	275	2 US-09-016-366A-17	Sequence 17, Appl
8	467	25.1	275	2 US-08-978-404B-12	Sequence 12, Appl
9	462	24.8	273	2 US-09-016-366A-19	Sequence 19, Appl
10	462	24.8	273	2 US-08-978-404B-14	Sequence 14, Appl
11	461	24.7	274	2 US-09-016-366A-21	Sequence 21, Appl
12	461	24.7	274	2 US-08-978-404B-16	Sequence 16, Appl
13	459	24.6	273	2 US-08-978-404B-3	Sequence 3, Appl
14	457	24.5	245	4 US-08-944-483-69	Sequence 69, Appl
15	456	24.5	249	4 US-09-079-970A-5	Sequence 5, Appl
16	456	24.5	267	2 US-09-016-366A-23	Sequence 23, Appl
17	456	24.5	267	2 US-08-978-404B-18	Sequence 18, Appl
18	456	24.5	273	2 US-08-978-404B-6	Sequence 6, Appl
19	455.5	24.4	492	4 US-09-342-749-2	Sequence 2, Appl
20	451	24.2	245	4 US-09-079-970A-6	Sequence 6, Appl
21	446.5	24.0	283	3 US-08-807-151-1	Sequence 1, Appl
22	446.5	24.0	638	3 US-08-681-151-3	Sequence 3, Appl
23	442.5	23.8	454	4 US-09-518-046-2	Sequence 2, Appl
24	440	23.6	238	4 US-08-944-483-64	Sequence 64, Appl
25	434.5	23.3	269	2 US-08-978-404B-10	Sequence 10, Appl
26	434	23.3	248	4 US-08-944-483-63	Sequence 63, Appl
27	424	22.8	435	4 US-09-008-271A-6	Sequence 6, Appl

28	408	21.9	416	2 US-09-000-846-2	Sequence 2, Appl
29	402.5	21.6	387	4 US-09-032-215-8	Sequence 8, Appl
30	402.5	21.6	387	4 US-09-032-215-13	Sequence 13, Appl
31	399.5	21.4	400	4 US-09-004-731-30	Sequence 30, Appl
32	399.5	21.4	400	4 US-09-004-731-33	Sequence 33, Appl
33	399.5	21.4	400	4 US-08-749-699-30	Sequence 30, Appl
34	399.5	21.4	400	4 US-08-749-699-33	Sequence 33, Appl
35	398	21.4	228	1 US-08-278-091-10	Sequence 10, Appl
36	398	21.4	228	1 US-08-483-859-10	Sequence 10, Appl
37	398	21.4	228	1 US-08-472-173-10	Sequence 10, Appl
38	398	21.4	228	2 US-08-487-167-10	Sequence 10, Appl
39	398	21.4	228	2 US-08-482-816-10	Sequence 10, Appl
40	398	21.4	228	2 US-08-296-149-10	Sequence 10, Appl
41	398	21.4	228	2 US-08-801-499-10	Sequence 10, Appl
42	398	21.4	228	2 US-08-615-271-10	Sequence 10, Appl
43	398	21.4	228	3 US-09-074-660-10	Sequence 10, Appl
44	398	21.4	228	3 US-09-074-659-10	Sequence 10, Appl
45	398	21.4	228	3 US-09-106-468-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1  
US-09-008-271A-3  
: Sequence 3, Application US/09008271A  
: Patent No. 6203979  
: GENERAL INFORMATION:  
: APPLICANT: Bandman, Olga  
: Hillman, Jennifer L.  
: Yue, Henry  
: Guegler, Karl J.  
: Corley, Neil C.  
: Tang, Tom Y.  
: Shah, Purvi  
: TITLE OF INVENTION: HUMAN PROTEASE MOLECULES  
: NUMBER OF SEQUENCES: 24  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Incyte Pharmaceuticals, Inc.  
: STREET: 3174 Porter Dr.  
: CITY: Palo Alto  
: STATE: CA  
: COUNTRY: USA  
: ZIP: 94304  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Diskette  
: OPERATING SYSTEM: DOS  
: SOFTWARE: FASTSEQ for Windows Version 2.0  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/008,271A  
: FILING DATE: 16-Jan-1998  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: <Unknown>  
: FILING DATE: <Unknown>  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Mohan-Peterson, Sheela  
: REGISTRATION NUMBER: 41,201  
: REFERENCE/DOCKET NUMBER: PF-0458 US  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 650-855-0555  
: TELEFAX: 650-845-4166  
: INFORMATION FOR SEQ ID NO: 3:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 314 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: IMMEDIATE SOURCE:  
: LIBRARY: PROSTAT03  
: CLONE: 789927  
: SEQUENCE DESCRIPTION: SEQ ID NO: 3 :  
US-09-008-271A-3

```

; TELEFAX: 847/938-2623
;
; TELEX:
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 299 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
US-08-944-483-66

```

Query Match 27.1%; Score 504; DB 4; Length 299;  
Best Local Similarity 39.0%; Pred. No. 4.le-44;  
Matches 113; Conservative 42; Mismatches 103; Indels 32; Gaps 7;

QY 28 VVGQDAAAGRPWQVSLHFDHNFIYGGSLVSERLILTAAHCIQTPTWTFTSYVMGSIIT 87  
DB 1 ITGSSAVAGQPWQVSIITYEGHVCGGSLSVEQVLSAHCFFSEHHKEAYEVLGAHQ 60  
QY 88 VGDSRRKRYKYYSK-RVIHPKY--QDTTADVALLKSSQVTTSAILPICLPSPVTKQLAI 144  
DB 61 LDSYSEDAKVSTLDIIPHSYLQEGSQGDIALLOLSRPTITSYIRPCLPAANASFPN 120  
QY 145 PPGCWTVTGKGVKESDRDYHSALQEAEPPIIDROACEOLYNPINIGIFLPALEP-VIKEDK 203  
DB 121 GLHCTVTGWGHVAPSVLLTPKPLQOQLEVPLISRETNCNLYN---IDAKPEEPHFVQEDM 177  
QY 204 ICAGDTQNMKDCSGDGGLSCHIDGVWIQTGVVSWGLECG-KSLPGVYTNVIYYQKWI 262  
DB 178 VCAGYVEGGKDACQGSGGPLSCPVEGLWLTYGIWSWGACCARRNPVGYTLASSYASI 237  
QY 263 NATISR-----ANNLDFSD-----FLFPVILLSLAL 288  
DB 238 QSKVTELPQPRVPQTOEQSDPSNLCSHLAFSSAPQAQGLLRPLFLPLGL 287

RESULT 3  
US-09-016-366A-15  
Sequence 15, Application US/09016366A  
Patent No. 5955431  
GENERAL INFORMATION:  
APPLICANT: Stevens, Richard L.  
APPLICANT: Huang, Chifu  
TITLE OF INVENTION: FAST CELL PROTEASE PEPTIDE  
TITLE OF INVENTION: INHIBITORS  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02210-2211  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016,366A  
FILING DATE: January 30, 1998  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/037,090  
FILING DATE: 05-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Plumer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7093  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441

Query Match 28.2%; Score 526; DB 4; Length 314;  
Best Local Similarity 38.1%; Pred. No. 2.3e-46;  
Matches 122; Conservative 53; Mismatches 109; Indels 36; Gaps 9;

QY 1 MGPAGCAFTLLLLGISV-----CCQPYSSRVVGGQDAAGRWPMQVSLH 46  
DB 1 MGARGALLALLARAGLRKPESQEAAPLSGPCGRVVITSRIVGGEDAELGRWPWGSLR 60  
QY 47 FDHNFYGGSLVSERLILTAHAHCIOPTWTFTS-----YVWLGSITVCDG-----RKRVKY 97  
DB 61 LWDSHVCVGSLLSHRWALTAAHCFE-TYSDLSDPGSMVQFGQUTSMPSFWSQLQAYTRY 119  
QY 98 YVSKIIVHPKY-QDTTADVALLKLSSOVTFTSAILPICLPSPVTKQLAIPPCFCWVTGNGKV 156  
DB 120 FVSNYLSPRYLGNSPDYIALVKLSAPVYTKHKTOPICLOASFEFENRDWCWTGNGYI 179  
QY 157 KESSDRDYHSALQEAEPPIIDROACEOLYNPINIGIFLPALEPVIKEDIKACGDTQNKKDSC 216  
DB 180 KEDEALPSPHLTQEQQVAIIINNSCNHLF-----LKYSFRKDIEGDMVCAGNAQGGKDAC 234  
QY 217 KDSSGGLSCHIDGVWIQTGVVSWGLECGK-SLPGVYTNVIYYQKWINAISRNANLDFS 275  
DB 235 FGDSGGLACNKNGLYQIGVYVSWGVCGRNRPVGTNTISHHEFIWKLMQAQ-SGM SQP 293  
QY 276 DFLFPI---VLLSLALLCP 291  
DB 294 DPSWPLLFPLWALPLLPG 313

RESULT 2  
US-08-944-483-66  
Sequence 66, Application US/08944483  
Patent No. 6232456  
GENERAL INFORMATION:  
APPLICANT: COHEN, MAURICE  
APPLICANT: COLPITTS, TRACEY L.  
APPLICANT: FRIEDMAN, PAULA N.  
APPLICANT: GRANADOS, EDWARD N.  
APPLICANT: KLASS, MICHAEL R.  
APPLICANT: RUSSELL, JOHN C.  
APPLICANT: STEWART, KENT D.  
APPLICANT: STROUPE, STEVEN D.  
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS  
TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
TITLE OF INVENTION: OF THE PROSTATE  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,483  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6183.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847/935-1729



TELEX:  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 276 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-016-366A-15

Query Match 25.8%; Score 480; DB 2; Length 276;  
Best Local Similarity 37.8%; Pred. No. 1.1e-41;  
Matches 104; Conservative 51; Mismatches 88; Indels 32; Gaps 9;  
QY 10 LLLLGISVCGQPVYSS-----RVVGGODAAAGRPQVSLHFDHNF---IYGGSLV 58  
Db 6 LLLLWALSLSLASVYSAPRANRQVIGVGGHEASESKRPQVSLRFLKLNWIHFCCGSLI 65  
QY 59 SERLILTAACHCIQIP-TWTTFSTYVLMGSIIVGDSRKRVKY-----VSKIVIHPKY--Q 109  
Db 66 HPQWVLTAACHVGHPIKSPQLFRVQL-----REQYLYGGDQLLSLNRIVWHPHYTTA 117  
QY 110 DTTADVALLKSSQVTTSAIPICLPSTVTKQLAIPPCWVTGKVKKESSDRDYHSALQ 169  
Db 118 EGGADVALLEEVNVTSTHPIISLPASETFPPGTSCWVTGWDIDNDEPLPPYPLK 177  
QY 170 EAEVPIIDROACEOLYNPIGIFLPALPEVKEKICAGDTQNMKDSCKGDSGGPLSCHID 229  
Db 178 QVKVPIVENSICDRKYH-TGLYTGDDFPVHDGMLCAGNTR--RDSGCGSGGGLVCKVK 234  
QY 230 GWTQGTGVWSMGLECGK-SLPGVYTNVIYQKWIN 263  
Db 235 GTWLQAGVWSWEGCAQPNKPGIYTRYTYLDWIH 269

RESULT 4  
US-08-978-404B-21  
; Sequence 21, Application US/08978404B  
; Patent No. 5968782  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
; TITLE OF INVENTION: FIBRINOGEN  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,404B  
; FILING DATE: 25-NOV-97  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/032,354  
; FILING DATE: 04-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7090  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 21:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 276 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5968782e  
US-08-978-404B-21  
Query Match 25.8%; Score 480; DB 2; Length 276;  
Best Local Similarity 37.8%; Pred. No. 1.1e-41;  
Matches 104; Conservative 51; Mismatches 88; Indels 32; Gaps 9;  
QY 10 LLLLGISVCGQPVYSS-----RVVGGODAAAGRPQVSLHFDHNF---IYGGSLV 58  
Db 6 LLLLWALSLSLASVYSAPRANRQVIGVGGHEASESKRPQVSLRFLKLNWIHFCCGSLI 65  
QY 59 SERLILTAACHCIQIP-TWTTFSTYVLMGSIIVGDSRKRVKY-----VSKIVIHPKY--Q 109  
Db 66 HPQWVLTAACHVGHPIKSPQLFRVQL-----REQYLYGGDQLLSLNRIVWHPHYTTA 117  
QY 110 DTTADVALLKSSQVTTSAIPICLPSTVTKQLAIPPCWVTGKVKKESSDRDYHSALQ 169  
Db 118 EGGADVALLEEVNVTSTHPIISLPASETFPPGTSCWVTGWDIDNDEPLPPYPLK 177  
QY 170 EAEVPIIDROACEOLYNPIGIFLPALPEVKEKICAGDTQNMKDSCKGDSGGPLSCHID 229  
Db 178 QVKVPIVENSICDRKYH-TGLYTGDDFPVHDGMLCAGNTR--RDSGCGSGGGLVCKVK 234  
QY 230 GWTQGTGVWSMGLECGK-SLPGVYTNVIYQKWIN 263  
Db 235 GTWLQAGVWSWEGCAQPNKPGIYTRYTYLDWIH 269

RESULT 5  
US-08-978-404B-8  
; Sequence 8, Application US/08978404B  
; Patent No. 5968782  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
; TITLE OF INVENTION: FIBRINOGEN  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,404B  
; FILING DATE: 25-NOV-97  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/032,354  
; FILING DATE: 04-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7090  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 270 amino acids

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5968782e
US-08-978-404B-8

Query Match 25.7%; Score 479.5; DB 2; Length 270;
Best Local Similarity 35.2%; Pred. No. 1.2e-41;
Matches 101; Conservative 51; Mismatches 84; Indels 51; Gaps 10;

QY 11 LLLLGISVCGQPVYSSR-----VVGQDAAAGRPQVSLRPHDNF---IYGGSL 57
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 4 LLLAL-----PLFSLMHRSPLCQEWLVGGQEPAGNKWPMQVSLRANEYWRHFCGSL 58
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 58 VSERLILTAAHCIQPTTFTSYVWLGSIYVGD-----RKRVKY-----VSKIVI 104
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 59 IHPQWLVTAACHVGP-----TIADPNKVRVQLRKQVLYYHDLHLAVSRIT 104
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 105 HPKYQDQT--ADVALLKSSQVTFTSAILPTCLPSVTQKLAIPFCWVTGKVKYESSDR 162
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 105 HPTFYATQNGADIALLELKNPNVNISSVHPVSLPPASETFPSGLCWVTGNGNDVSL 164
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 163 DYHSALQAEVPIIDRQACEQLYNPIGIFLPALEPVIKEDKICAGDTQNMKDSCKGSGG 222
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 165 PPPPLKEVQVPVENQLCDLKYHK-GVYTGDNTHIVRDDMLCAGNEGH--DSCQDGG 221
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 223 PLSCHDGVMIQTGVVSNGLGEG-KSLPGVYTNVYIYOKWINAIRS 268
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 222 PLVCKVNGTWLQAGVYVWEGGALPNRFGIYTRVYIYLDWIHRVYVK 268
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 6
US-08-978-404B-5
; Sequence 5, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 274 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5968782e
US-08-978-404B-8

Query Match 25.1%; Score 468; DB 2; Length 274;
Best Local Similarity 36.6%; Pred. No. 1.9e-40;
Matches 102; Conservative 53; Mismatches 84; Indels 40; Gaps 10;

QY 10 LLLLGISVCGQPVYSS-----RVVGGQDAAAGRPQVSLRPHDNF---IYGGSLV 58
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 4 LLLALLSLASLVAAPCPKQVRGIVGGREASBSPQVSLRKFSEFWMHFCGSLI 63
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 59 SERLILTAAHCI-----OPTTFTSYVWLGSIYVGDSCRKRKY-----VSKIVHPK 107
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 64 HPQWLVTAACHVGLHKSP-----LFRVQL-----REQLYYADQLLTNRVTVVHPH 111
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 108 YQ--DPTADVALLKSSQVTFTSAILPTCLPSVTQKLAIPFCWVTGKVKYESSDRDYH 165
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 112 YTVEDGADIALLELIPVNVSTHPIISLPASETFPSGTSCTWGTGMDIDSDPELPP 171
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 166 SALQAEVPIIDRQACEQLYNPIGIFLPALEPVIKEDKICAGDTQNMKDSCKGSGGSL 225
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 172 YPLQKVKPIVENSLCDKRYH-TGLYTGDDVDIVQDGLCAGNTR--DSCQDGGSLV 228
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 226 CHIDGVMIQTGVVSNGLGEG-KSLPGVYTNVYIYOKWIN 263
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 229 CKVKGWTLQAGVYVWEGGCAEANRPGIYTRVYIYLDWIH 267
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 7
US-09-016-366A-17
; Sequence 17, Application US/09016366A
; Patent No. 5955431
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Huang, Chifu
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
; TITLE OF INVENTION: INHIBITORS
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,366A
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/037,090
; FILING DATE: 05-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7093
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5968782e
US-08-978-404B-5

```

MOLECULE TYPE: protein  
US-09-016-366A-17

Query Match 25.1%; Score 467; DB 2; Length 275;  
Best Local Similarity 37.2%; Pred. No. 2.5e-40;  
Matches 103; Conservative 50; Mismatches 88; Indels 36; Gaps 11;  
QY 11 LLLLGISVCGQPYS-----SRVGGGDAAGRWPMQVSLHFDHNF---IYGGSL 57  
DB 4 LLLALPVLASRAYAAPVQALQQAGIVGGQAPRSKWPQVSLRVRDYMWHFCGSL 63  
QY 58 VSRLLILTAHCLQPTWTFSTYVWLGSTVGDSTRKRVY-----VSKIVLHPKYQ-- 109  
DB 64 IHPQWLTAHCLGPDVKD-----LATLRV-QLREQHLYYQDQLLPVSRIVHPQFYII 116  
QY 110 DTTADVALLKLSQVFTTSAILPCLPSTVKQLAIPPF--CWTGKGKVKESSDRDYHSA 167  
DB 117 QTGADIALLELEPVSNISSRVHTVMLPPASE--TFFPGMPCWVTGWDVNDDEPLPPFP 174  
QY 168 LQAEVPIIDROACEOLYNPIGIFLPALEVIKEDKICAGDTQNMKDSCKGDSGGPLSCH 227  
DB 175 LKQVKVPIMENHICDAKYH-LGAYTGDDVRIIRDMLCAGNSQ--RDSCKGDSGGPLVCK 231  
QY 228 IDGVWITQTVGWSGLECGK-SLPGVTNTVYYQKWIN 263  
DB 232 VNGTWLQAGVWSWDEGCAQPNRPGIYTRVYYLDWIH 268

RESULT 8  
US-08-978-404B-12  
; Sequence 12, Application US/08978404B  
; Patent No. 5968782  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
; TITLE OF INVENTION: FIBRINOGEN  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,404B  
; FILING DATE: 25-NOV-97  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/032,354  
; FILING DATE: 04-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7090  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 275 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5968782  
US-08-978-404B-12

Query Match 25.1%; Score 467; DB 2; Length 275;  
Best Local Similarity 37.2%; Pred. No. 2.5e-40;  
Matches 103; Conservative 50; Mismatches 88; Indels 36; Gaps 11;  
QY 11 LLLLGISVCGQPYS-----SRVGGGDAAGRWPMQVSLHFDHNF---IYGGSL 57  
DB 4 LLLALPVLASRAYAAPVQALQQAGIVGGQAPRSKWPQVSLRVRDYMWHFCGSL 63  
QY 58 VSRLLILTAHCLQPTWTFSTYVWLGSTVGDSTRKRVY-----VSKIVLHPKYQ-- 109  
DB 64 IHPQWLTAHCLGPDVKD-----LATLRV-QLREQHLYYQDQLLPVSRIVHPQFYII 116  
QY 110 DTTADVALLKLSQVFTTSAILPCLPSTVKQLAIPPF--CWTGKGKVKESSDRDYHSA 167  
DB 117 QTGADIALLELEPVSNISSRVHTVMLPPASE--TFFPGMPCWVTGWDVNDDEPLPPFP 174  
QY 168 LQAEVPIIDROACEOLYNPIGIFLPALEVIKEDKICAGDTQNMKDSCKGDSGGPLSCH 227  
DB 175 LKQVKVPIMENHICDAKYH-LGAYTGDDVRIIRDMLCAGNSQ--RDSCKGDSGGPLVCK 231  
QY 228 IDGVWITQTVGWSGLECGK-SLPGVTNTVYYQKWIN 263  
DB 232 VNGTWLQAGVWSWDEGCAQPNRPGIYTRVYYLDWIH 268

RESULT 9  
US-09-016-366A-19  
; Sequence 19, Application US/09016366A  
; Patent No. 5955431  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; APPLICANT: Huang, Chifu  
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
; TITLE OF INVENTION: INHIBITORS  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,366A  
; FILING DATE: January 30, 1998  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/037,090  
; FILING DATE: 05-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7093  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 273 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-016-366A-19



QY 11 LLLLGISVC-----GQPYSSRVVGGQDAAAGRWPMQVSL-----HFDHNF I 52  
 Db 3 LLLALPVLASRAYAAPAQALQVRGIVGGQAPRSKWPQVSLRVHGPYWMHF----- 57  
 QY 53 YGSLYSERLILTAACICQPTWTFSTYVWLGSIYVGDSTRKVKY-----YSKIVIH 106  
 Db 58 CGSLHPQVWLVAACVGPDKD-----LAALRV-QLREQHLYYQDQLLPVSRIVHP 110  
 QY 107 KYDQTT--ADVALLKSSQVTFSSAILPICLPSVTQKLAIPPF--CWVWGKVKSSDR 162  
 Db 111 QFTAQIGADIALLEEPKVVSSHVHTVLPASE--TFPPGMPCWVTGWDVNDRL 168  
 QY 163 DYHSALQAEVPIIDQACQOLNPICIFLPALEPVYKEDKICAGDTQNMKDSKGDGSG 222  
 Db 169 PPFPPLQKVRPMENHICDAKYH-LGAYTGDVRIYRDDMLCAGNTR--RDSCQDGG 225  
 QY 223 PLSCHDGWIQGVVSWGLECGK-SLPGYVTNVIYKWIN 263  
 Db 226 PLVCKVNGTWLQAGVSWGEGCAQPNRPGIYTRTYLDWIH 267

RESULT 12  
 US-08-978-404B-16  
 ; Sequence 16, Application US/08978404B  
 ; Patent No. 5968782  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Stevens, Richard L.  
 ; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
 ; NUMBER OF SEQUENCES: 74  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
 ; STREET: 600 Atlantic Avenue  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02210-2211  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/978,404B  
 ; FILING DATE: 25-NOV-97  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/032,354  
 ; FILING DATE: 04-DEC-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Plumer, Elizabeth R.  
 ; REGISTRATION NUMBER: 36,637  
 ; REFERENCE/DOCKET NUMBER: B0801/7090  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-720-3500  
 ; TELEFAX: 617-720-2441  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 16:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 274 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: No. 5968782e  
 ; US-08-978-404B-16

Query Match .24.7%; Score 461; DB 2; Length 274;  
 Best Local Similarity 36.9%; Pred. No. 1e-39;  
 Matches 104; Conservative 50; Mismatches 82; Indels 46; Gaps 12;  
 QY 11 LLLLGISVC-----GQPYSSRVVGGQDAAAGRWPMQVSL-----HFDHNF I 52  
 Db 3 LLLALPVLASRAYAAPAQALQVRGIVGGQAPRSKWPQVSLRVHGPYWMHF----- 57

Db 3 LLLALPVLASRAYAAPAQALQVRGIVGGQAPRSKWPQVSLRVHGPYWMHF----- 57  
 QY 53 YGSLYSERLILTAACICQPTWTFSTYVWLGSIYVGDSTRKVKY-----YSKIVIH 106  
 Db 58 CGSLHPQVWLVAACVGPDKD-----LAALRV-QLREQHLYYQDQLLPVSRIVHP 110  
 QY 107 KYDQTT--ADVALLKSSQVTFSSAILPICLPSVTQKLAIPPF--CWVWGKVKSSDR 162  
 Db 111 QFTAQIGADIALLEEPKVVSSHVHTVLPASE--TFPPGMPCWVTGWDVNDRL 168  
 QY 163 DYHSALQAEVPIIDQACQOLNPICIFLPALEPVYKEDKICAGDTQNMKDSKGDGSG 222  
 Db 169 PPFPPLQKVRPMENHICDAKYH-LGAYTGDVRIYRDDMLCAGNTR--RDSCQDGG 225  
 QY 223 PLSCHDGWIQGVVSWGLECGK-SLPGYVTNVIYKWIN 263  
 Db 226 PLVCKVNGTWLQAGVSWGEGCAQPNRPGIYTRTYLDWIH 267

RESULT 13  
 US-08-978-404B-3  
 ; Sequence 3, Application US/08978404B  
 ; Patent No. 5968782  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Stevens, Richard L.  
 ; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
 ; NUMBER OF SEQUENCES: 74  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
 ; STREET: 600 Atlantic Avenue  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02210-2211  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/978,404B  
 ; FILING DATE: 25-NOV-97  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/032,354  
 ; FILING DATE: 04-DEC-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Plumer, Elizabeth R.  
 ; REGISTRATION NUMBER: 36,637  
 ; REFERENCE/DOCKET NUMBER: B0801/7090  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-720-3500  
 ; TELEFAX: 617-720-2441  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 273 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: No. 5968782e  
 ; US-08-978-404B-3

Query Match 24.6%; Score 459; DB 2; Length 273;  
 Best Local Similarity 35.9%; Pred. No. 1.6e-39;  
 Matches 101; Conservative 49; Mismatches 85; Indels 46; Gaps 10;  
 QY 11 LLLLGISVCQPYSS-----RVVGGQDAAAGRWPMQVSLHFDHNF---IYGSLSVS 59  
 Db 4 LLLALPVLASRAYAAPAQALQVRGIVGGQAPRSKWPQVSLRVHGPYWMHF----- 63



Search completed: November 22, 2001, 02:38:27  
Job time: 2626 sec

**This Page Blank (uspto)**



Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being predicted and is derived by analysis of the total score distribution.

score greater than or equal to the score of the result being predicted. NO is the number of results predicted by chance to have a score greater than or equal to the score of the result being predicted. NO is derived by analysis of the total score distribution.

## SUMMARIES

PT activity and viability - particularly the testis-specific protease  
PT HELA2 used for modulation of fertility and as tumour suppressor  
XX  
PS Example 14; Fig 18A; 167pp; English.

XX The present sequence represents the amino acid sequence of murine HELA2.  
CC Human HELA2 was isolated from HeLa cells. HELA2 has high homology to  
CC serine proteases. The protein is involved in or associated with  
CC regulation of cell activity and/or viability. Administration of  
CC recombinant HELA2 (also called testisin) is used to increase fertility.  
CC Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of  
CC testicular germ cell cancers (seminoma) and is also expressed in some  
CC non-testicular cancers (of colon, pancreas, prostate and ovary), so is  
CC a marker/potential therapeutic target for cancer. The promoter from the  
CC HELA2 gene is useful for testis-specific expression of other genes,  
CC e.g. for gene therapy or modulation of fertility. Drugs that block  
CC activity of HELA2 should have antitumour activity (other than in  
CC testis) while in testis recombinant HELA2 should stop growth of tumours  
CC and normalise sperm development (eliminating the need for orchidectomy).  
CC Identification of mutant forms of HELA2 can be used to diagnose  
CC infertility.

XX Sequence 285 AA;

Query Match 29.1%; Score 541.5; DB 19; Length 285;  
Best Local Similarity 41.8%; Pred. No. 2.1e-40;  
Matches 119; Conservative 40; Mismatches 109; Indels 17; Gaps 6;  
QY 19 CGQPVYSSRVVGGQDAAGRPWQVSLHFDHNPFIYGGSLVSERLIILTAACHIQPTWTTF 78  
DB 7 cghrtipsrvggdgaelgrwpqgsrlwvghlcatllnrrwvtaahcfkqkndpfd 66  
QY 79 YTVWLGSITVGDSSRKRY-----KYVYSKIVHPKYQDT-TADVALLKLSQVTFTSAILP 132  
DB 67 wtvgfelterpalwnlqasnyryqiediflskpyeqpndialklspvtynnfigp 126  
QY 133 ICLPSVTKQLAIPFPFCWVTGCGWKSSDRDYHSALQAEVPIIDRQACBOLYNPIGIFL 192  
DB 127 icllnstykfenrtcdwvfgwaigedeslpsntlqevqvalinnsmcnhmykk----- 181  
QY 193 PALEPVTKEDKICAGDTQNMKDSKGGPLSCHIDGVWIOGTGVVSWGLECGK-SLPGV 251  
DB 182 pdftrnlwgmvcagtppegkdacfgdsgplacdgdvtvwygvvswgigcgrprpvg 241  
QY 252 YTNVYVYQKWINATISRANLDFSLFPVLLSLA----LLCPS 292  
DB 242 ytnishhynwigtmir-ngllrpdvpvlliflilawassllrpa 285

## RESULT 2

AAW77297  
ID AAW77297 standard; Protein; 314 AA.

AC AAW77297;

XX 07-JAN-1999 (first entry)

XX Amino acid sequence of long isoform of HELA2.

XX Serine protease; regulation; cell activity; viability; HELA2; ATC2;  
KW BCOM3; testisin; fertility; suppressor; testicular germ cell cancer;  
KW seminoma; testis-specific expression; antitumour; sperm development;  
KW infertility.

XX Homo sapiens.

XX WO9836054-A1.

XX 20-AUG-1998.

XX 13-FEB-1998; 98WO-AU00085.

XX

PR 18-NOV-1997; 97AU-0000422.  
XX 13-FEB-1997; 97AU-0005101.  
PA (AMRA-) AMRAD OPERATIONS PTY LTD.

XX Antalis TM, Hooper JD;

XX WPI; 1998-480768/41.

DR N-PSDB; AAV59119.

XX New serine protease(s) and kinase involved in regulating cell  
PT activity and viability - particularly the testis-specific protease  
PT HELA2 used for modulation of fertility and as tumour suppressor

XX Claim 3; Pages 62-64; 167pp; English.

XX The present sequence represents the amino acid sequence of the long  
CC isoform of HELA2. cDNA generated from HeLa cells and PAI-2 expressing  
CC HeLa cells was amplified using PCR primers AAV48312-13. Three new  
CC sequences were detected in the 480 bp amplicon. These sequences are  
CC designated HELA2 and ATC2 which have high homology to serine proteases  
CC and BCOM3 which has homology to a kinase. The proteins are involved in  
CC or associated with regulation of cell activity and/or viability.  
CC Administration of recombinant HELA2 (also called testisin) is used to  
CC increase fertility. Downregulation of HELA2 reduces fertility. HELA2 is  
CC also a suppressor of testicular germ cell cancers (seminoma) and is also  
CC expressed in some non-testicular cancers (of colon, pancreas, prostate  
CC and ovary), so is a marker/potential therapeutic target for cancer. The  
CC promoter from the HELA2 gene is useful for testis-specific expression of  
CC other genes, e.g. for gene therapy or modulation of fertility. Drugs  
CC that block activity of HELA2 should have antitumour activity (other than  
CC in testis) while in testis recombinant HELA2 should stop growth of  
CC tumours and normalise sperm development (eliminating the need for  
CC orchidectomy). Identification of mutant forms of HELA2 can be used to  
CC diagnose infertility.

XX Sequence 314 AA;

Query Match 28.5%; Score 531; DB 19; Length 314;  
Best Local Similarity 38.4%; Pred. No. 2e-39;  
Matches 123; Conservative 52; Mismatches 109; Indels 36; Gaps 9;

QY 1 MGPAGCAFTLLLLLGISV-----CGQPVYSSRVVGGQDAAGRPWQVSLH 46

DB 1 mgargallalllaraglrkpesgeaplsqpcgrvritsriyvggedaelgrwpqgsrlr 60

QY 47 FDHNFYGGSLVSERLIILTAACHIQPTWTTFs----YTVWLGSITVGDs-----RKRKY 97

DB 61 lwdshvcgvsllshrwaltahcfe-tydsldspgwmvqfgltsmpsfswlqayvtry 119

QY 98 YVSKIVHPKY-ODTTADVALLKLSQVTFTSAILPICLPSVTKQLAIPFPFCWVTGWKV 156

DB 120 fvnilylsprylngnsydlavklspvtytkhlpclqastfefeirtcdwvtgwy1 179

QY 157 KESSDRDYHSALQAEVPIIDRQACBOLYNPIGIFLPALEPVTKEDKICAGDTQNMKDS 216

DB 180 kedealpsphltqevqvaiinnsmcnhlf-----lkysfrkdifgdmvcagnagdgkdc 234

QY 217 KGSGGPLSCHIDGVWIOGTGVVSWGLECGK-SLPGVYTNVYVYQKWINATISRANLDFS 275

DB 235 fgdsggplacnkdglyqvgvswgvcgrprpvgvtytnishhfwigkimaq-sgmsqp 293

QY 276 DELFPI---VLLSLALLCP 291

DB 294 dpswplllffllwalpllgp 313

## RESULT 3

AAW77296  
ID AAW77296 standard; Protein; 314 AA.

XX AAW77296;

XX 07-JAN-1999 (first entry)  
XX Amino acid sequence of the short isoform of HELA2.  
DE Serine protease; regulation; cell activity; viability; HELA2; ATC2;  
KW BCOM3; testisin; fertility; suppressor; testicular germ cell cancer;  
KW seminoma; testis-specific expression; antitumour; sperm development;  
KW infertility.  
XX Homo sapiens.  
XX W09836054-A1.  
XX 20-AUG-1998.  
XX 13-FEB-1998; 98WO-AU00085.  
XX 18-NOV-1997; 97AU-0000422.  
PR 13-FEB-1997; 97AU-0005101.  
XX (AMRA-) AMRAD OPERATIONS PTY LTD.  
XX Antalis TM, Hooper JD;  
PI WPI; 1998-480768/41.  
DR N-PSDB; AAV59118.  
XX New serine protease(s) and kinase involved in regulating cell  
PT activity and viability - particularly the testis-specific protease  
PT HELA2 used for modulation of fertility and as tumour suppressor  
XX  
PS Claim 2; Fig 6; 167pp; English.  
XX  
CC The present sequence represents the amino acid sequence of the short  
CC isoform of HELA2. cDNA generated from HeLa cells and PAI-2 expressing  
CC HeLa cells was amplified using PCR primers AAV48312-13. Three new  
CC sequences were detected in the 480 bp amplicon. These sequences are  
CC designated HELA2 and ATC2 which have high homology to serine proteases  
CC and BCOM3 which has homology to a kinase. The proteins are involved in  
CC or associated with regulation of cell activity and/or viability.  
CC Administration of recombinant HELA2 (also called testisin) is used to  
CC increase fertility. Downregulation of HELA2 reduces fertility. HELA2 is  
CC also a suppressor of testicular germ cell cancers (seminoma) and is also  
CC expressed in some non-testicular cancers (of colon, pancreas, prostate  
CC and ovary), so is a marker/potential therapeutic target for cancer. The  
CC promoter from the HELA2 gene is useful for testis-specific expression of  
CC other genes, e.g. for gene therapy or modulation of fertility. Drugs  
CC that block activity of HELA2 should have antitumour activity (other than  
CC in testis) while in testis recombinant HELA2 should stop growth of  
CC tumours and normalise sperm development (eliminating the need for  
CC orchidectomy). Identification of mutant forms of HELA2 can be used to  
CC diagnose infertility.  
XX  
SQ Sequence 314 AA;

Query Match 28.2%; Score 526; DB 19; Length 314;  
Best Local Similarity 38.1%; Pred. No. 5.6e-39;  
Matches 122; Conservative 53; Mismatches 109; Indels 36; Gaps 9;

QY 1 MGPAGCAFTULLLLGIVS-----CGQPYSSRVVGGQDAAGRPWQVSLH 46  
DB 1 mgargallalliaraglrkpesgeapigpcgrvritsriggedaelgrwpwgsrlr 60  
QY 47 FDNHNYGSLSERLLTAHAIQPTWTTF-----YTWLGSITVGD-----RRKVKY 97  
DB 61 lwdshvcgvsllshrvaltaahcfe-tysdlsqpsgmvmvqfgqltmspsfwsiqayvtry 119  
QY 98 YVSKIVHPKY-QDITADVALKLSQVTFSTAILPLCLPSVTKQLAIPPCVWTGKGV 156  
DB 120 fvsniylsprlgnspdyialvklspavtytkhlpicicqastfefeentdcwvtgwgy 179

QY 157 KESSDRDYHSALQAEVPIIDRQACEQLYNPIGIFLPALEPVIKEDKICAGDTQNMKDC 216  
DB 180 kedealpsphltqevqvaiinnsmcnhlf-----lkysfrkdlfgdmvcagnagggkdac 234  
QY 217 KDSGGPLSCHIDGVWITQTGVSWGLECGK-SLPGVYTNVYIYQKWINATISRANNIDFS 275  
DB 235 fgsdggplacnkglyqigvswgvcgrpnrgvynishhfewiqkimag-sgmsqp 293  
QY 276 DELFPI---VLLSLALLCP 291  
DB 294 dpswpllffpllwpllpg 313  
RESULT 4  
AY06434  
ID RAY06434 standard; Protein; 314 AA.  
XX  
AC RAY06434;  
XX  
DT 27-SEP-1999 (first entry)  
XX  
DE Human protease HUPM-3.  
XX  
KW Protease; human; HUPM-3; cell proliferation; cancer;  
KW Immune disorder; inflammation; therapy.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..19  
FT /note= "putative signal peptide"  
FT Protein 20..314  
FT /note= "putative mature protein"  
FT Modified-site 39  
FT /note= "protein kinase C phosphorylation site"  
FT Modified-site 58  
FT /note= "protein kinase C phosphorylation site"  
FT Modified-site 73  
FT /note= "protein kinase C phosphorylation site"  
FT Active-site 82  
FT Modified-site 86  
FT /note= "casein kinase II phosphorylation site"  
FT Modified-site 127  
FT /note= "protein kinase C phosphorylation site"  
FT Modified-site 134  
FT /note= "casein kinase II phosphorylation site"  
FT Modified-site 161  
FT /note= "casein kinase II phosphorylation site"  
FT Modified-site 167  
FT /note= "N-glycosylated"  
FT Modified-site 190  
FT /note= "casein kinase II phosphorylation site"  
FT Modified-site 200  
FT /note= "N-glycosylated"  
FT Modified-site 212  
FT /note= "protein kinase C phosphorylation site"  
FT Active-site 238  
FT Modified-site 273  
FT /note= "N-glycosylated"  
FT Modified-site 291  
FT /note= "casein kinase II phosphorylation site"  
FT  
XX W09936550-A2.  
XX  
PD 22-JUL-1999.  
XX  
PF 12-JAN-1999; 99WO-US00655.  
XX  
PR 16-JAN-1998; 98US-0008271.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Bandnan O, Corley NC, Guegler KJ, Hillman JL, Shah P;

PI Tang YT, Yue H;  
XX WPI: 1999-430616/36.  
DR N-PSDB; AAX87151.  
XX  
XX Novel human protease molecules useful in the treatment of  
PT developmental disorders and/or cancers  
XX  
XX Claim 1: Page 71-72; 90pp; English.  
PS  
XX The present sequence represents novel human protease HUPM-3, as  
CC deduced from the consensus sequence (see AAX87151) of overlapping  
CC cDNA clones obtained from various libraries. Northern analysis  
CC shows expression of HUPM-3 in cardiovascular, haematopoietic and  
CC male reproductive cDNA libraries. Approximately 86% of these  
CC libraries are associated with neoplastic disorders. The invention  
CC provides 12 new human proteases, i.e. HUPM-1 to -12 (see  
CC AAY06432-43), and the polynucleotides encoding them (see AAX87149-60).  
CC Also provided are vectors, host cells and methods for producing  
CC HUPM polypeptides, as well as agonists and antagonists of HUPM.  
CC Methods for treating or preventing cell proliferative disorders  
CC and immune disorders using HUPM or HUPM antagonists are claimed.  
XX  
XX Sequence 314 AA;  
SQ  
  
Query Match 28.2%; Score 526; DB 20; Length 314;  
Best Local Similarity 38.1%; Pred. No. 5.6e-39;  
Matches 122; Conservative 53; Mismatches 109; Indels 36; Gaps 9;  
  
QY 1 MGPAGCAFTLLLLGISV-----CGQPVYSRVVGGQDAAGRWPMQVSLH 46  
Db 1 mgargallalllaraglrpesgeaaplgpcgrvritsriyvggedaelgrwpqgsrlr 60  
  
QY 47 FDHNFYGGSLVSRILTAHICIQPTWTF-----YTVWLGSITVGD-----RKRVKY 97  
Db 61 lwshvcgvallhrwaltaahcfe-tystdspsgvmvqfgiltsmpstfswlqayvtry 119  
  
QY 98 YVSKIVTHPKY-QDTADVALLKLSQVTFSAITLCLPSVKQLAIPFCWVTGHWKV 156  
Db 120 fvsnylsprylgnsydlalvksaptytkhlpclqastfedenrdcwvtgwyi 179  
  
QY 157 KESSDRVHSAQEAPEIIDROACEVLYPIGIFLPALEPIKEDKICAGDTQNMKDS 216  
Db 180 kedaelpsptlqevqvainmscnhlf-----lkystrkdfgdmvcagagggkdc 234  
  
QY 217 KGDSGGPLSCHIDGWITQTGVVSWGECGK-SLPGVYTNVYIYOKWKNATISRANND 275  
Db 235 fgdsggplacnknlgwyigvsvvgcgrpnrgvntnshhfwiklmaq-sgmsqp 293  
  
QY 276 DFLFPI----VLLSLALCP 291  
Db 294 dpswplliffllwalplgp 313  
  
RESULT 5  
AAY13388  
ID AAY13388 standard; Protein: 314 AA.  
XX AAY13388;  
AC  
XX  
XX 25-JUN-1999 (first entry)  
DT  
XX Amino acid sequence of protein PRO303.  
DE  
XX Secreted protein; transmembrane protein; human; enterocolitis;  
KW Zollinger-Ellison syndrome; gastrointestinal ulceration;  
KW congenital microvillus atrophy; skin disease; cell growth;  
KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;  
KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;  
KW fibromodulin; dermal scarring; Usher Syndrome; Atrophla areata;  
KW anti-thrombotic; wound healing; tissue repair.  
XX

OS Homo sapiens.  
XX W09914328-A2.  
XX  
XX PD 25-MAR-1999.  
XX  
XX PF 16-SEP-1998; 98WO-US19330.  
XX  
XX 25-NOV-1997; 97US-0066840.  
PR 17-SEP-1997; 97US-0059113.  
PR 17-SEP-1997; 97US-0059115.  
PR 17-SEP-1997; 97US-0059117.  
PR 17-SEP-1997; 97US-0059119.  
PR 17-SEP-1997; 97US-0059121.  
PR 17-SEP-1997; 97US-0059122.  
PR 17-SEP-1997; 97US-0059184.  
PR 18-SEP-1997; 97US-0059263.  
PR 18-SEP-1997; 97US-0059266.  
PR 15-OCT-1997; 97US-0062125.  
PR 17-OCT-1997; 97US-0062285.  
PR 17-OCT-1997; 97US-0062287.  
PR 21-OCT-1997; 97US-0063486.  
PR 24-OCT-1997; 97US-0062814.  
PR 24-OCT-1997; 97US-0062816.  
PR 24-OCT-1997; 97US-0063045.  
PR 24-OCT-1997; 97US-0063120.  
PR 24-OCT-1997; 97US-0063121.  
PR 24-OCT-1997; 97US-0063127.  
PR 24-OCT-1997; 97US-0063128.  
PR 27-OCT-1997; 97US-0063329.  
PR 27-OCT-1997; 97US-0063327.  
PR 28-OCT-1997; 97US-0063541.  
PR 28-OCT-1997; 97US-0063542.  
PR 28-OCT-1997; 97US-0063544.  
PR 28-OCT-1997; 97US-0063549.  
PR 28-OCT-1997; 97US-0063550.  
PR 28-OCT-1997; 97US-0063564.  
PR 29-OCT-1997; 97US-0063435.  
PR 29-OCT-1997; 97US-0063704.  
PR 29-OCT-1997; 97US-0063732.  
PR 29-OCT-1997; 97US-0063738.  
PR 29-OCT-1997; 97US-0063734.  
PR 29-OCT-1997; 97US-0064215.  
PR 29-OCT-1997; 97US-0063735.  
PR 31-OCT-1997; 97US-0063870.  
PR 31-OCT-1997; 97US-0064103.  
PR 03-NOV-1997; 97US-0064248.  
PR 07-NOV-1997; 97US-0064809.  
PR 12-NOV-1997; 97US-0065186.  
PR 17-NOV-1997; 97US-0065846.  
PR 18-NOV-1997; 97US-0065693.  
PR 21-NOV-1997; 97US-0066120.  
PR 21-NOV-1997; 97US-0066364.  
PR 24-NOV-1997; 97US-0066772.  
PR 24-NOV-1997; 97US-0066466.  
PR 24-NOV-1997; 97US-0066770.  
PR 24-NOV-1997; 97US-0066511.  
PR 24-NOV-1997; 97US-0066453.  
XX  
XX (GETH ) GENENTECH INC.  
XX  
XX Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;  
XX WPI: 1999-229533/19.  
XX N-PSDB; AAX52259.  
XX  
XX New isolated human genes and polypeptides used in, e.g. treatment of  
PT gastrointestinal ulceration  
XX  
XX Claim 12; Fig 92; 320pp; English.  
XX  
XX AAY13344-403 represent secreted and transmembrane human proteins.  
CC The cDNA sequences are obtained from cDNA libraries, prepared from  
CC

CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.  
 CC The encoded polypeptides have specific uses based on their homology to  
 CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders  
 CC associated with the preservation and maintenance of gastrointestinal  
 CC mucosa and the repair of acute and chronic mucosal lesions  
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal  
 CC ulceration and congenital microvillus atrophy), skin diseases associated  
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial  
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),  
 CC potent effects on cell growth and development, diseases related to  
 CC growth or survival of nerve cells including Parkinson's disease,  
 CC Alzheimer's disease, ALS, neuropathies or cancer. PRO285 can be used as  
 CC for fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used  
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment  
 CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an  
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have  
 CC therapeutic applications in wound healing and tissue repair; PRO317 can  
 CC be used for treating problems of the kidney, uterus, endometrium, blood  
 CC vessels, or related tissue, e.g. in the heart of genital tract.  
 XX  
 SQ Sequence 314 AA;

Query Match 28.2%; Score 526; DB 20; Length 314;  
 Best Local Similarity 38.1%; Pred. No. 5.6e-39;  
 Matches 122; Conservative 53; Mismatches 109; Indels 36; Gaps 9;

QY 1 MGPAGCAFTLLLLGISV-----CGQPVYSSRVVGGODAAAGRWQVSLH 46  
 Db 1 mgargallalllaraglrkpesqaaplsqpcrrvitsrivggedaelgrwpwqgsrlr 60  
 QY 47 FDHNFYVGGSLVSRERLTLTAACIQPTWTTFSS-----YTVMLGSIYVGDSS-----RKRVKY 97  
 Db 61 lwshvcgvslshrwaltaahcfe-tysdlsdpsgmvmvqfgltsmpsfwslyqayvtry 119  
 QY 98 YVSKIVTHPKY-QDFTADVALLKSSQVTFSAIPLCLPSVTKQLAIPFCWWTGNGKV 156  
 Db 120 fvnilylspylgnspsydlavklisapvtytkhlpicldqastfefeentdcwvtgwygi 179  
 QY 157 KESSDRDYHSALEAEVPIIDROACEOLYNPIGIFLPALEPVIKEDKICAGDTQNMKDCS 216  
 Db 180 kedealpshtqlqvvaainnsmcnhlf-----lkysfrkdfgdmvcagngagqgdac 234  
 QY 217 KDGSGGPLSCHIDGWIQTQVSWGLECGK-SLPGVYTNVYIYQKWINATISRANNLDFS 275  
 Db 235 fgdsggplacnkglyqigvswgvcgrpnrgpytnishhfwiklmaq-sgmsqp 293  
 QY 276 DFLFPI-----VLLSLALLCP 291  
 Db 294 dpswpllffllwalpllpg 313

RESULT 6  
 AAW97116  
 ID AAW97116 standard; Protein; 314 AA.  
 XX  
 AC AAW97116;  
 XX  
 DT 04-MAY-1999 (first entry)  
 XX  
 DE A human eosinophil serine protease.  
 XX  
 KW Human; eosinophil; serine protease; allergic disease; infectious disease;  
 KW tumour; granulomatous disease; collagen disease; vascular inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN JP11032768-A.  
 XX  
 PD 09-FEB-1999.  
 XX  
 PF 16-JUL-1997; 97JP-0191319.  
 XX

PR 16-JUL-1997; 97JP-0191319.  
 XX (ONOV ) ONO PHARM CO LTD.  
 PA  
 XX WPI; 1999-183825/16.  
 DR N-PSDB; AAX15336.  
 XX  
 PT New eosinophils serine protease - useful for prevention and  
 PT treatment of allergic, infectious, tumour, granulomatous and collagen  
 PT diseases  
 XX  
 PS Claim 1; Page 9-10; 18pp; Japanese.  
 XX  
 CC The present sequence represents a human eosinophil serine protease.  
 CC The protease is useful in drug compositions for the prevention and  
 CC treatment of allergic diseases, infectious diseases, tumour diseases,  
 CC granulomatous diseases, collagen diseases and vascular inflammation.  
 XX  
 SQ Sequence 314 AA;

Query Match 28.2%; Score 526; DB 20; Length 314;  
 Best Local Similarity 38.1%; Pred. No. 5.6e-39;  
 Matches 122; Conservative 53; Mismatches 109; Indels 36; Gaps 9;

QY 1 MGPAGCAFTLLLLGISV-----CGQPVYSSRVVGGODAAAGRWQVSLH 46  
 Db 1 mgargallalllaraglrkpesqaaplsqpcrrvitsrivggedaelgrwpwqgsrlr 60  
 QY 47 FDHNFYVGGSLVSRERLTLTAACIQPTWTTFSS-----YTVMLGSIYVGDSS-----RKRVKY 97  
 Db 61 lwshvcgvslshrwaltaahcfe-tysdlsdpsgmvmvqfgltsmpsfwslyqayvtry 119  
 QY 98 YVSKIVTHPKY-QDFTADVALLKSSQVTFSAIPLCLPSVTKQLAIPFCWWTGNGKV 156  
 Db 120 fvnilylspylgnspsydlavklisapvtytkhlpicldqastfefeentdcwvtgwygi 179  
 QY 157 KESSDRDYHSALEAEVPIIDROACEOLYNPIGIFLPALEPVIKEDKICAGDTQNMKDCS 216  
 Db 180 kedealpshtqlqvvaainnsmcnhlf-----lkysfrkdfgdmvcagngagqgdac 234  
 QY 217 KDGSGGPLSCHIDGWIQTQVSWGLECGK-SLPGVYTNVYIYQKWINATISRANNLDFS 275  
 Db 235 fgdsggplacnkglyqigvswgvcgrpnrgpytnishhfwiklmaq-sgmsqp 293  
 QY 276 DFLFPI-----VLLSLALLCP 291  
 Db 294 dpswpllffllwalpllpg 313

RESULT 7  
 AAB12132  
 ID AAB12132 standard; Protein; 314 AA.  
 XX  
 AC AAB12132;  
 XX  
 DT 02-FEB-2001 (first entry)  
 XX  
 DE Hydrophobic domain protein from clone HP03116 isolated from KB cells.  
 XX  
 KW Human; secreted protein; membrane protein; hydrophobic domain;  
 KW proliferation control; differentiation induction; material transport;  
 KW biophylaxis; signal receptor; ion channel; transporter; immunostimulant;  
 KW immunosuppressant; haematopoiesis regulator; chemotactic; chemokinetic;  
 KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;  
 KW autoimmune disease; Alzheimer's disease; Parkinson's disease; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200029448-A2.  
 XX  
 PD 25-MAY-2000.  
 XX

```
PF 17-NOV-1999; 99WO-JP06412.
XX
XX 17-NOV-1998; 98JP-0326255.
PR 22-DEC-1998; 98JP-0364315.
PR 16-MAR-1999; 99JP-0069811.
PR 27-APR-1999; 99JP-0119299.
PR 19-MAY-1999; 99JP-0138169.
XX
XX (SAGA ) SAGAMI CHEM RES CENT.
PA (PROF-) PROTEGENE INC.
XX
XX Kato S, Kimura T;
PI
XX WPI: 2000-387753/33.
DR N-PSDB; AAA62005, AAA62015.
XX
XX Proteins comprising hydrophobic regions, such as secretory and membrane
PT proteins, useful in research and diagnostics and having various
PT activities e.g. immunomodulatory, antiinflammatory, chemokinetic,
PT hemostatic, thrombolytic -
XX
XX Claim 1; Page 238-240; 410pp; English.
PS
XX Secretory proteins play important roles in the proliferation control, the
XX differentiation induction, the material transport and the biophylaxis of
CC cells. Membrane proteins have important roles as signal receptors, ion
CC channels and transporters. The present sequence is a human protein which
CC has at least one hydrophobic domain. This protein may be a secretory or a
CC membrane protein. The present protein may have cytokine and cell
CC proliferation/differentiation activity, immune stimulating or suppressing
CC activity, haematopoiesis activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, anti-inflammatory activity and tumour
CC inhibition activity. The present protein could therefore be used for
CC treatment of autoimmune disease, Alzheimer's disease, Parkinson's
XX disease, and cancer.
XX
XX Sequence 314 AA;

Query Match 28.2%; Score 526; DB 21; Length 314;
Best Local Similarity 38.1%; Pred. No. 5.6e-39;
Matches 122; Conservative 53; Mismatches 109; Indels 36; Gaps 9;

QY 1 MGPAGCAFTLLLLGISV-----CGQPVYSRRVVGGDAAAAGRPWPQVSLH 46
Db 1 mgargallalllaraglrpesqaaplsgpcgrrrvitsrivvgdaelgrwpwqgsrlr 60
QY 47 FDHNFYXGSLVSRILITAAHCITQPTWTF-----YTVMLGSTIVGDS-----RKRVKY 97
Db 61 lwdshvcgvsllshrwaltaahcfelcysdlpsgvmvqfgqitpsfswlsqayvtry 119
QY 98 YVSKIVTHPKY-QDTHADVALLKLSQVTFSTALPCLPSVTKQLAIPPCWVTGKGK 156
Db 120 fvsniysprylgnsydyialvksapvtyckhiqpiclqastfedenrdtcwvtgwyi 179
QY 157 KESSRDYHSAQEAPEIIDRQACEOLYNPIGIFLPALEPVIKEDICAGDTONMKDSC 216
Db 180 kedaelpshtqlcvqvaillnsmcnhlf-----lksyfrkdfgdmvcnagggkdac 234
QY 217 KGDSGGPLSCHIDGVIQITGVVSWGLBEGK-SLPGVVTNTVIYQKWINATISRANNLD 275
Db 235 fgdgggplacnknlgwqigvsvvgcgrprnrgvntnshhfwiklmaq-sgmsqp 293
QY 276 DFLPPI-----VLLSLALCP 291
Db 294 dpswplllffllwalpllgp 313

RESULT 8
AAB80256
ID AAB80256 standard; Protein; 314 AA.
XX
```

```
AC AAB80256;
XX
XX 24-APR-2001 (first entry)
XX
XX Human PRO303 protein.
XX
XX Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;
KW antiparkinsonian nootropic; neuroprotective; vulnerary; cardiant;
KW antiangiogenic; vasotropic; antiasthmatic; antirheumatic; cancer;
KW antiarthritic; antifertility; antidiabetic; antiviral; diabetes;
KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
KW ischaemia; inflammation.
XX
XX Homo sapiens.
XX
XX WO200104311-A1.
XX
XX 18-JAN-2001.
XX
XX 22-FEB-2000; 2000WO-US04414.
XX
XX 07-JUL-1999; 99US-0143048.
PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 08-SEP-1999; 99WO-US20594.
PR 13-SEP-1999; 99WO-US20944.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 29-NOV-1999; 99WO-US28214.
PR 30-NOV-1999; 99WO-US28313.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 99WO-US30999.
XX
XX (GETH ) GENENTECH INC.
PA
XX
XX Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
PI Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kljavin IJ;
PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
PI Williams PM, Wood WI;
XX
XX WPI: 2001-081051/09.
DR N-PSDB; AAF72417.
XX
XX Sixty one nucleic acids encoding PRO polypeptides which are useful in
PT the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
PT squamous cell carcinoma) and neurodegenerative diseases (e.g.
PT Alzheimer's disease) -
XX
XX Claim 1; Fig 92; 393pp; English.
PS
XX The present sequence is one of sixty one novel secreted and
CC transmembrane PRO polypeptides. The PRO polypeptides are
CC useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
CC squamous cell carcinoma), gastrointestinal disorders (e.g.
CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.
CC endometrial bleeding, angiogenesis, ischaemias such as coronary
CC ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,
CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and
CC diabetes and retinal disorders such as retinitis pigmentosum.
CC The PRO nucleic acids have applications in molecular biology, including
CC use as hybridization probes, and in chromosome and gene mapping.
XX
XX Sequence 314 AA;

Query Match 28.2%; Score 526; DB 22; Length 314;
Best Local Similarity 38.1%; Pred. No. 5.6e-39;
Matches 122; Conservative 53; Mismatches 109; Indels 36; Gaps 9;
```

QY	1	MG	PAGCAFTLLLLGTSV-----CGQPVYSRVVGGODAAAGRPWQVSLH	46
Db	1	mg	argallalllaraglrkpesqaaplsqpcgrvritsrivggedaelgrwpwgslr	60
QY	47	FD	HNFIYGSLSYSERLILTAACIQPTWTFS-----YTVMLGSITVGD	97
Db	61	ld	shvcgvsllshrwaltaahcfe-tysdlsdpqsgvmvqfgqltsmpsfwslqaytry	119
QY	98	YV	KIVIHPKY-QDTTADVALLKLSQVFTSAILPICLPSTVKQLAIPPFVWTG	156
Db	120	fv	snlylspriygnspydialvklsapvtytkhiqpiclqastfentdcwvtgwyi	179
QY	157	KES	DRDYHSALQAEVPIIDROACBQLYNPIGIFLPALEPVIKEDKICAGDTQ	216
Db	180	ke	dealpshtiqevqvalinnsmchlf-----lkysfrkdifgdmvcagna	234
QY	217	KD	SGGLSCHIDGWIQTGVVSWGLECGK-SLPGYVTNVYIYQKWINATISR	275
Db	235	fd	sggplacnknlgwylqigvsvgvcgrprpgrvytnishhfwiklmaq-s	293
QY	276	DF	LEPFI-----VLLSLALLCP	291
Db	294	dp	swpllfpllwlpplgp	313
RESULT	9			
ID	AA	Y91871	standard; Protein; 327 AA.	
XX	AC	AA	Y91871;	
XX	DT	19-JUL-2000	(first entry)	
XX	DE	Human	cancer-specific gene protein, Pro104.	
XX	KW	Human	cancer-specific gene protein; Pro104; serine protease; diagnosis;	
XX	KW	staging; imaging; treating;	gynecologic cancer; testicular cancer.	
XX	OS	Homo	sapiens.	
FH	Key	Location/Qualifiers		
FT	Domain	54..58		
FT	/note=	"conserved serine protease motif"		
FT	Cleavage-site	54..55		
FT	Domain	55..327		
FT	/label=	active_protease_domain		
PN	WO	200016805-A1.		
XX	PD	30-MAR-2000.		
XX	PF	23-SEP-1999;	99WO-US21774.	
XX	PR	23-SEP-1998;	98US-0101522.	
XX	PA	(DIAD-)	DIADEXUS LLC.	
XX	PI	Ali SM, Cafferkey R;		
XX	DR	WPI; 2000-283453/24.		
XX	DR	N-PSDB; AAA08505.		
PT	Methods	for diagnosing, staging, imaging and treating gynecologic and		
PT	testicular	cancers by measuring expression of a cancer specific gene -		
XX	Example	2; Page 32-33; 36pp; English.		
XX	This	protein sequence is encoded by human cancer-specific gene (CSG),		
CC	Pro104.	Pro104 is a serine protease that shares 31 percent homology with		
CC	human	hepsin at the amino acid level. When aligned with other serine		
CC	proteases,	Pro104 shares all the conserved amino acid motifs that are		
CC	characteristic	of all other serine proteases, in particular a highly		
CC	conserved	RIVG3 sequence (see AAY91872). The inventive methods comprise		
CC	measuring	expression of CSG in order to diagnose, stage, image and treat		
CC	gynecologic	and testicular cancers. The levels of CSG are compared to		
CC	levels	found in a normal human control (a change in the measured level		
CC	of	CSG is associated with the presence of the cancer). The early		
CC	diagnosis	of cancers improves the success rate of therapeutic protocols.		
XX	Sequence	327 AA;		
Query	Match	28.2%; Score 526; DB 21; Length 327;		
Best	Local	Similarity 38.1%; Pred. No. 5.9e-39;		
Matches	122; Conservative	53; Mismatches 109; Indels 36; Gaps 9;		
QY	1	MG	PAGCAFTLLLLGTSV-----CGQPVYSRVVGGODAAAGRPWQVSLH	46
Db	14	mg	argallalllaraglrkpesqaaplsqpcgrvritsrivggedaelgrwpwgslr	73
QY	47	FD	HNFIYGSLSYSERLILTAACIQPTWTFS-----YTVMLGSITVGD	97
Db	74	ld	shvcgvsllshrwaltaahcfe-tysdlsdpqsgvmvqfgqltsmpsfwslqaytry	132
QY	98	YV	KIVIHPKY-QDTTADVALLKLSQVFTSAILPICLPSTVKQLAIPPFVWTG	156
Db	133	fv	snlylspriygnspydialvklsapvtytkhiqpiclqastfentdcwvtgwyi	192
QY	157	KES	DRDYHSALQAEVPIIDROACBQLYNPIGIFLPALEPVIKEDKICAGDTQ	216
Db	193	ke	dealpshtiqevqvalinnsmchlf-----lkysfrkdifgdmvcagna	247
QY	217	KD	SGGLSCHIDGWIQTGVVSWGLECGK-SLPGYVTNVYIYQKWINATISR	275
Db	248	fd	sggplacnknlgwylqigvsvgvcgrprpgrvytnishhfwiklmaq-s	306
QY	276	DF	LEPFI-----VLLSLALLCP	291
Db	307	dp	swpllfpllwlpplgp	326
RESULT	10			
ID	AA	Y73388	standard; Protein; 290 AA.	
XX	AC	AA	Y73388;	
XX	DT	24-FEB-2000	(first entry)	
XX	DE	HTRM	clone 3376404 protein sequence.	
XX	KW	HTRM;	human transcriptional regulatory molecule; arteriosclerosis; AIDS;	
XX	KW	arteriosclerosis; cirrhosis;	cancer; leukaemia; diabetes mellitus;	
XX	KW	Addison's	disease; multiple sclerosis; rheumatoid arthritis; infection;	
XX	OS	Homo	sapiens.	
XX	PN	WO	9957144-A2.	
XX	PD	11-NOV-1999.		
XX	PF	04-MAY-1999;	99WO-US09935.	
XX	PR	05-MAY-1998;	98US-0084254.	
XX	PR	07-AUG-1998;	98US-0095827.	
XX	PR	02-OCT-1998;	98US-0102745.	
XX	PA	(INCY-)	INCYTE PHARM INC.	
XX	PI	Hillman	JL, Bandman O, Lal P, Yue H, Reddy R, Tang YT;	
XX	PI	Gerstin	EH, Patterson C, Baughn MR, Azimzai Y, Lu DAM;	
XX	DR	WPI; 2000-052941/04.		
XX	DR	N-PSDB; AAZ52473.		

XX New peptides useful for diagnosis, prevention and treatment of cancer  
PT and immune disorders  
XX  
XX  
PS Claim 1; Page 145-146; 193pp; English.

XX AAY73325-773389 are human transcriptional regulator molecule (HTRM)  
CC protein sequences. The HTRM protein and nucleotide sequences are useful  
CC for preventing or treating disorders associated with decreased expression  
CC or activity of HTRM which include cell proliferative disorders such as  
CC arteriosclerosis and cirrhosis; cancers including adenocarcinoma and  
CC leukemia; immune disorders such as AIDS, Addison's disease, diabetes  
CC mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus  
CC erythematosus, and myasthenia gravis; infections and trauma. Antagonists  
CC of the HTRM polypeptides are useful for treating or preventing disorders  
CC associated with increased expression or activity of HTRMs. HTRM  
CC polypeptides, their immunogenic fragments or oligopeptides are useful  
CC for screening libraries of compounds in drug screening techniques.  
CC Polynucleotides encoding HTRM are useful for blocking the transcription  
CC of mRNA and regulating gene function by modulating the activity of  
CC HTRM. Vectors expressing HTRM or agonists can also be used to prevent or  
CC treat disorder associated with decreased HTRM expression. Antibodies  
CC which specifically bind HTRM and polynucleotides encoding HTRM are useful  
CC for diagnosing disorders associated with the expression of HTRM,  
CC particularly in assays that detect the expression of HTRM. Nucleotide  
CC sequences encoding HTRM may be useful to generate hybridization probes  
CC useful in mapping the naturally occurring genomic sequence and to detect  
CC differences in gene sequences among normal, carrier and affected  
CC individuals. Using diagnostic assays, cancer can be detected prior to  
CC the appearance of clinical symptoms and thereby progression of cancer can  
CC be prevented by aggressive treatment or preventive measures.

XX  
XX Sequence 290 AA;

Query Match 27.8%; Score 517.5; DB 21; Length 290;  
Best Local Similarity 41.2%; Pred. NO. 2.9e-38;  
Matches 114; Conservative 38; Mismatches 112; Indels 13; Gaps 6;

QY 3 PACGAFTLLLLG-----ISVCGQPVSRRVGGQDAAAGRWPMQVSLHFDHFIYGG 56  
DB 4 PAAPVLLLLLCFSGRAKAATACGRPMRMVGGQDQEGEPWQVSIQRNGSHFCGGS 63

QY 57 LVSERLILTAHCICQPTWTFTSYTVMWLGSTVGDSSRRVKY-YVSKIVIHPKYQDT--TA 113  
DB 64 LIAEQWLTAAHCFRNTSETSLYVLLGARQLVGPQHMYRVQVESNPPLYGGTASSA 123

QY 114 DVALLKLSQVTFSTALPCLPSVTQKLAIPPCVWTGKVKESDRDYHSALEAEV 173  
DB 124 DVALVEAPVPFTNYILPCLPDPVSIFETGMNCWVTGWSPEEDLLPEPRILKLV 183

QY 174 PTDROACEOLYNPIGIFLPALEP-VIKEDKICAGDTQNMKDSCKGSGPLSCHIDGVW 232  
DB 184 PIDTPKCNLLYSKDTF--GYQPKTKNDMLCAGFEQKKDACKGSGGPLVCLVGGSW 241

QY 233 IQTGWSWGLECGK-SLPGVYTNVYVYQKINATISR 268  
DB 242 LQAGVISGEGCARQNPVYIRVTAHNHWIHLIPK 278

RESULT 11  
AAB73945  
ID AAB73945 standard; Protein; 290 AA.  
XX  
AC AAB73945;  
XX  
XX 29-MAY-2001 (first entry)  
XX Human protease T.  
XX Human; protease T; serine protease; dermatological; desquamation;  
KW skin care; laundry; detergent; shampoo; skin flaking.  
XX

OS Homo sapiens.  
XX  
PN WO200116293-A2.  
XX  
PD 08-MAR-2001.  
XX  
XX 30-AUG-2000; 2000WO-US23823.  
XX  
XX 31-AUG-1999; 99US-0386653.  
XX  
XX (ORTH ) ORTHO-MCNEIL PHARM INC.  
XX  
XX Darrow AL, Qi J, Andrade-Gordon P;  
XX WPI: 2001-265889/27.  
XX N-PSDB; AAF76994.  
XX  
XX New serine protease termed protease T, useful for treating and  
PT preventing skin flaking or imbalance of desquamation -  
XX  
XX Claim 11; Fig 1; 83pp; English.

XX The present sequence is human protease T protein. The protein is useful  
CC for treating a condition mediated by protease T. It is useful for  
CC treating an imbalance of desquamation, by topical application of  
CC a pharmaceutical composition containing the protease. The composition is  
CC useful as a topical skin care composition. It is useful as a laundry  
CC detergent, shampoo, hard surface cleaning composition, and dish care  
CC cleaning composition. Protease T protein is useful for treating and  
CC preventing skin flaking. It is less immunogenic to sensitive individuals  
CC and it provides efficient proteolytic activity in a non-natural  
CC environment.

XX  
XX Sequence 290 AA;

Query Match 27.8%; Score 517.5; DB 22; Length 290;  
Best Local Similarity 41.2%; Pred. NO. 2.9e-38;  
Matches 114; Conservative 38; Mismatches 112; Indels 13; Gaps 6;

QY 3 PACGAFTLLLLG-----ISVCGQPVSRRVGGQDAAAGRWPMQVSLHFDHFIYGG 56  
DB 4 PAAPVLLLLLCFSGRAKAATACGRPMRMVGGQDQEGEPWQVSIQRNGSHFCGGS 63

QY 57 LVSERLILTAHCICQPTWTFTSYTVMWLGSTVGDSSRRVKY-YVSKIVIHPKYQDT--TA 113  
DB 64 LIAEQWLTAAHCFRNTSETSLYVLLGARQLVGPQHMYRVQVESNPPLYGGTASSA 123

QY 114 DVALLKLSQVTFSTALPCLPSVTQKLAIPPCVWTGKVKESDRDYHSALEAEV 173  
DB 124 DVALVEAPVPFTNYILPCLPDPVSIFETGMNCWVTGWSPEEDLLPEPRILKLV 183

QY 174 PTDROACEOLYNPIGIFLPALEP-VIKEDKICAGDTQNMKDSCKGSGPLSCHIDGVW 232  
DB 184 PIDTPKCNLLYSKDTF--GYQPKTKNDMLCAGFEQKKDACKGSGGPLVCLVGGSW 241

QY 233 IQTGWSWGLECGK-SLPGVYTNVYVYQKINATISR 268  
DB 242 LQAGVISGEGCARQNPVYIRVTAHNHWIHLIPK 278

RESULT 12  
AAB36481  
ID AAB36481 standard; Protein; 319 AA.  
XX  
AC AAB36481;  
XX  
XX 02-MAR-2001 (first entry)  
XX Fusion gene with human serine protease catalytic domain protein #12.  
XX Activation construct; catalytic; fusion gene; expression vector;  
KW proteolysis; serine protease; zymogen precursor; characterisation;



analysis; modulator; identification.

XX Homo sapiens.

OS Synthetic.

XX WO200066709-A2.

XX 09-NOV-2000.

XX 13-APR-2000; 2000WO-US09973.

XX 30-APR-1999; 99US-0303162.

XX (ORTH ) ORTHO-MCNEIL PHARM RES INC.

XX Darrow A, Qi J, Andrade-Gordon P;

XX WPI; 2000-687533/67.

XX Expression vector for producing recombinantly producing serine protease domains, comprising a presequence, a prosequence, and a cloning site for the insertion of catalytic domain cassette -

XX Disclosure; Page 47-50; 89pp; English.

XX The present invention describes an expression vector (I) comprising in frame and in order, a presequence, a prosequence, and a cloning site for the in frame insertion of catalytic domain cassette. (I) can be used as a modulator of proteins expressed from a zymogen activation construct. The recombinant catalytic domain of serine protease is useful for identifying compounds modulating the activity of proteases is expressed and activated from the zymogen activation construct. A method from the present invention comprises combining a modulator of the recombinant catalytic domain of a protease and measuring an effect of the modulator on the protein preferably inhibiting or enhancing its enzymatic activity or stimulation or inhibition of proteolysis mediated by the expressed catalytic domain. The present sequence represents a fusion gene with a human serine protease catalytic domain protein sequence from the present invention.

XX Sequence 319 AA;

Query Match 27.2%; Score 507; DB 21; Length 319;

Best Local Similarity 40.2%; Pred. No. 2.8e-37;

Matches 115; Conservative 41; Mismatches 102; Indels 28; Gaps 8;

QY 7 AFTLL-LLGISV-CGQPVY-----SSRVVGODAAAGRPQVSLH 46

Db 2 afdllscwallgtfgcgvdpdkdddaaalaapfddddkivgylaegqwpqvsit 61

QY 47 FDHNFYIGSLYSERILTAACHIQPTWTFSTVWLGSIYVWLSGDSRKRKYTVSK-IVIH 105

Db 62 yegvhvcgslvseqvsaahcfpsehkhayevklgahqidsydsedakvstlkdiiph 121

QY 106 PKY-ODTTADVALLKLSQVFTSAILPCLPSVTKQLAIPPCFWTGWGKVESDRD 163

Db 122 psylqsggdiallqlsrptfryirpicipaanasfnglhctvtgwhvapsvsl 181

QY 164 YHSAQEAEPVIDRQACBQLNPIGIFLPALEP-VIKEDKICAGDTQNMKSCRGDSG 222

Db 182 tpbqlqglevplisretcncln---ldakpeephfvqdmvcagvggkdcacqds 238

QY 223 PLSCHTDGVMIQTGVVSWGLEG-KSLPGVYTNVIYQKWINATIS 267

Db 239 plscpveglwylgtvswgdacgarnrpvytllassyaswiqskvt 284

RESULT 13

ID AAB67541

XX AAB67541 standard; Protein; 319 AA.

AC AAB67541;

XX 29-MAY-2001 (first entry)

XX Amino acid sequence of catalytic domain in CFEK2-6XHIS-TAG.

XX Expression vector; zymogen precursor; serine protease; prostatic; protease; inflammation; reproduction; epidermal tissue; skin care; neurological tissue; laundry detergent; stain-removing solution; prolactin; protease EK.

XX Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..22 /note= "chymotrypsinogen presequence"

FT Peptide 23..29 /note= "FLAG sequence"

FT Peptide 30..50 /note= "prosequence recognised by EK2"

FT Peptide 50..313 /note= "prostatic"

FT Peptide 314..319 /note= "His tag"

XX WO200116289-A2.

XX 08-MAR-2001.

XX 14-AUG-2000; 2000WO-US22283.

XX 31-AUG-1999; 99US-0386642.

XX (ORTH ) ORTHO-MCNEIL PHARM INC.

XX Darrow A, Qi J, Andrade-Gordon P;

XX WPI; 2001-218523/22.

XX N-PSDB; AAF55268.

XX An expression vector for the expression of inactive zymogen proteases useful for therapeutic or commercial products comprises a pre-sequence, a pro-sequence and a cloning site for in frame insertion of a catalytic domain cassette -

XX Claim 9; Fig 4A-D; 175pp; English.

XX The specification describes an expression vector system that will permit, through limited proteolysis, the activation of expressed zymogen precursors of serine proteases (e.g. prostatic) in a highly controlled and reproducible fashion. The expression vector comprises, in frame and in order, a pre-sequence, a pro-sequence and a cloning site for in frame insertion of a catalytic domain cassette. The expression vectors of the invention are useful for the expression of heterologous inactive zymogen proteases that can subsequently be proteolytically processed to generate the active enzyme product. The active enzyme product can be useful for directly treating diseases associated with inflammatory, reproductive, epidermal or neurological tissue or for identifying modulators of protease activity which can be used for treatment. The proteases can also be used in commercial products, e.g. laundry detergents, stain-removing solutions and skin care products. The present sequence is encoded by the catalytic domain in an expression vector of the invention. The construct encodes a prostatic protease sequence.

XX Sequence 319 AA;

Query Match 27.2%; Score 507; DB 22; Length 319;

Best Local Similarity 40.2%; Pred. No. 2.8e-37;

Matches 115; Conservative 41; Mismatches 102; Indels 28; Gaps 8;

QY 7 AFTLL-LLGISV-CGQPVY-----SSRVVGODAAAGRPQVSLH 46

Db 2 afdllscwallgtfgcgvdpdkdddaaalaapfddddkivgylaegqwpqvsit 61

QY	47	FDHNFTYGGSLVSRILILPAACHCIQPTWTFESYTVWLGSITVGDSSKRKYVYVSK-IVTH 105
DB	62	yegvhcvgsgslvseqwlsaaahcfpsehkhkeayevklghahqidsyedakvstlkdliph 121
QY	106	PKY-QDTADVALLKSSQVFTTSAILPICLPSTVKQLAIPFCWVTGWGKVKESSDRD 163
DB	122	psylqesgqgdialqlsrpittfsrlyrpicipaanasfpnglhtctvtgwhvapsvll 181
QY	164	YHSALQEAEPPIIDROACBOLYNPICIFLPALEP-VIKEDKICAGDTQNMKDSCKGDSGG 222
DB	182	tpkplqglevplisretcnclyn---idakpeephfvqedmvcagvyeggdacgds93 238
QY	223	PLSCHDGVMIQTGVVSWGLECG-KSLPGVNTVYIQKWINATIS 267
DB	239	plscpveglwyltgivswgdacgarmpgvytlassyaaswigskvt 284
RESULT 14		
ID	AAW77304	
AC	AAW77304 standard; Protein; 297 AA.	
XX	AAW77304;	
DT	07-JAN-1999	(first entry)
DE	Amino acid sequence of SP003LA, a homologue of HELA2.	
XX		
KW	Serrine protease; regulation; cell activity; viability; HELA2; ATC2;	
KW	BOOM3; testis; fertility; suppressor; testicular germ cell cancer;	
KW	seminoma; testis-specific expression; antitumour; sperm development;	
KW	infertility; human; chromosome 16p13.3.	
XX		
OS	Homo sapiens.	
FH	Key	
FT	Disulfide-bond 1	Location/Qualifiers
FT	/note= "likely to be involved in disulphide bonding"	
FT	Disulfide-bond 36	
FT	/note= "likely to be involved in disulphide bonding"	
FT	Misc-difference 51	
FT	/note= "likely to be a catalytic residue"	
FT	Disulfide-bond 52	
FT	/note= "likely to be involved in disulphide bonding"	
FT	Misc-difference 100	
FT	/note= "likely to be a catalytic residue"	
FT	Disulfide-bond 134	
FT	/note= "likely to be involved in disulphide bonding"	
FT	Disulfide-bond 167	
FT	/note= "likely to be involved in disulphide bonding"	
FT	Disulfide-bond 190	
FT	/note= "likely to be involved in disulphide bonding"	
FT	Disulfide-bond 201	
FT	/note= "likely to be involved in disulphide bonding"	
FT	Misc-difference 205	
FT	/note= "likely to be a catalytic residue"	
FT	Disulfide-bond 211	
FT	/note= "likely to be involved in disulphide bonding"	
FT	Disulfide-bond 229	
FT	/note= "likely to be involved in disulphide bonding"	
XX		
PN	WO9836054-A1.	
XX		
PD	20-AUG-1998.	
XX		
PF	13-FEB-1998;	98WO-AU00085.
XX		
PR	18-NOV-1997;	97AU-0000422.
PR	13-FEB-1997;	97AU-0005101.
XX		
PA	(AMRA-) AMRAD OPERATIONS PTY LTD.	
PI	Antalis TM, Hooper JD;	

FT Disulfide-bond 1 /note= "likely to be involved in disulphide bonding"  
FT Disulfide-bond 37 /note= "likely to be involved in disulphide bonding"  
FT Disulfide-bond 52 /note= "likely to be involved in disulphide bonding"  
FT Misc-difference 53 /note= "likely to be a catalytic residue"  
FT Disulfide-bond 53 /note= "likely to be involved in disulphide bonding"  
FT Misc-difference 101 /note= "likely to be a catalytic residue"  
FT Disulfide-bond 120 /note= "likely to be involved in disulphide bonding"  
FT Disulfide-bond 133 /note= "likely to be involved in disulphide bonding"  
FT Disulfide-bond 164 /note= "likely to be involved in disulphide bonding"  
FT Disulfide-bond 198 /note= "likely to be involved in disulphide bonding"  
FT Misc-difference 202 /note= "likely to be a catalytic residue"  
FT Misc-difference 208 /note= "likely to be a catalytic residue"  
FT Disulfide-bond 226 /note= "likely to be involved in disulphide bonding"

W09836054-A1.

20-AUG-1998.

13-FEB-1998; 98WO-AU000085.

18-NOV-1997; 97AU-0000422.

13-FEB-1997; 97AU-0005101.

(AMRA-) AMRAD OPERATIONS PTY LTD.

Antalis TM, Hooper JD;

WPI: 1998-480768/41.

N-PSDB; AAV59135.

New serine protease(s) and kinase involved in regulating cell activity and viability - particularly the testis-specific protease HELA2 used for modulation of fertility and as tumour suppressor

Example 15; Fig 20B; 167pp; English.

AAW7302-04 represent HELA2 homologues. The genes are found in a cluster on chromosome 16p13.3. HELA2 was isolated from HeLa cells, and has homology to serine proteases. The protein is involved in or associated with regulation of cell activity and/or viability. Administration of recombinant HELA2 (also called testisin) is used to increase fertility. Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of testicular germ cell cancers (seminoma) and is also expressed in some non-testicular cancers (of colon, pancreas, prostate and ovary), so is a marker/potential therapeutic target for cancer. The promoter from the HELA2 gene is useful for testis-specific expression of other genes, e.g. for gene therapy or modulation of fertility. Drugs that block activity of HELA2 should have antitumour activity (other than in testis) while in testis recombinant HELA2 should stop growth of tumours and normalise sperm development (eliminating the need for orchidectomy). Identification of mutant forms of HELA2 can be used to diagnose infertility.

Sequence 289 AA;

Query Match 26.7%; Score 497; DB 19; Length 289;

Best Local Similarity 38.8%; Pred. No. 1.9e-36;

Matches 112; Conservative 40; Mismatches 115; Indels 22; Gaps 8;

OY 19 CGQPVYSSRWGGQDAAGRWQVSLHFDH-NFIYGGSLVSRILITAAHCICQPTWTF 77

|| : ||||| ||||| | : |||| : ||||| :

Db 1 cghskeagrivgqdtqegrwpqvglwltsvghvcgsslihrwvltahcflrsedpg 60  
QY 78 SYTWLGSITVGDSSRRK-VKYYYSKIVIHPKYQDTN--ADVALLKLSQVTFTSAILPIC 134  
Db 61 lyhvkvggitpslsephsalvavrllvhsyhggttsgdalmeidsplq-asqfspic 119  
QY 135 LPSVTQQLAIPTPPFCWVTGKVKRESSDRDYHSALQAEVPIIDRQACEQLYNPIGIFLPA 194  
Db 120 lppqptplaigtvcwvnglg---vhsgealasvlqevavplldsnmcelmyh-lgepsla 175  
QY 195 LEPVIKEDKICAGDTQNMKDSCKGSGPLSCHIDGWIQTGVVSWGLECGKSL-PGVYT 253  
Db 176 gqrliqddmlcagsvggkdkscgsggplvcplndtwiqagivsvgfgcarfrpgvyt 235  
QY 254 NVIYYOKWINATISRAN-----NLDFSDFLFPVILSLALLC 290  
Db 236 qvisytdwigrtiaeshsgmgarpgapshsgtsrshpvllellitvc 284

Search completed: November 22, 2001, 02:37:43  
Job time: 2722 sec

---

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic -- nucleic search, using sw model

Run on: November 21, 2001, 23:30:45 ; Search time 1182.46 Seconds  
(without alignments)  
8321.990 Million cell updates/sec

Title: US-09-854-844-1  
Perfect score: 1041  
Sequence: 1 atgggacctgtgctgtgc.....ttattacaattgaaatga 1041

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 10228115 seqs, 4726426750 residues 20456230  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est33:\*  
26: gb\_est34:\*  
27: gb\_est35:\*  
28: gb\_est36:\*  
29: gb\_est37:\*  
30: gb\_est38:\*  
31: gb\_est39:\*  
32: gb\_est40:\*  
33: em\_estba:\*  
34: em\_estfun:\*  
35: em\_esthum1:\*  
36: em\_esthum2:\*  
37: em\_esthum3:\*  
38: em\_esthum4:\*  
39: em\_esthum5:\*  
40: em\_esthum6:\*  
41: em\_esthum7:\*  
42: em\_esthum8:\*  
43: em\_esthum9:\*  
44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est41:\*  
111: gb\_est42:\*  
112: gb\_est43:\*  
113: gb\_est44:\*  
114: gb\_est45:\*  
115: gb\_est46:\*  
116: gb\_est47:\*

No function

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	134.2	12.9	566	236	AQ987586
2	131	12.6	1050	192	AK006271 Mus muscu
3	130.2	12.5	975	150	BF582351 602101183
4	129.4	12.4	472	238	A2077203 RPCI-23-4
5	117	11.2	876	108	AL555870 AL555870
6	113.6	10.9	545	4	AA266251 mz49609.r
c 7	113	10.9	936	106	AL578261 AL578261
c 8	109	10.5	633	16	A1126185 qd81h05.x
9	107.2	10.3	479	7	AA448095 zw82c04.r
10	105.4	10.1	1089	192	AK016509 Mus muscu
11	94.2	9.0	833	150	BF582140 602099730
12	93	8.9	999	106	AL551470 AL551470
13	90.2	8.7	1256	192	AK007173 Mus muscu
14	89.4	8.6	1629	192	AK010640 Mus muscu
c 15	89.2	8.6	542	18	AI326140 mr64q08.x
c 16	85.6	8.2	551	120	AW765781 da77b09.y
17	83.8	8.0	560	4	AA266056 mz52d12.r
18	83	8.0	563	120	AW766877 da61b12.y
19	81.6	7.8	1006	169	BF789214 602105109
20	80.2	7.7	1323	192	AK014645 Mus muscu
21	79.6	7.6	442	7	AA401525 zu63q08.r
22	79.6	7.6	443	7	AA401517 zu63q08.r
23	79.6	7.6	454	7	AA400045 zu62h08.r
c 24	78	7.5	412	7	AA448166 zw83a10.s
c 25	78	7.5	480	7	AA400074 zu70h11.r
26	77.2	7.4	507	7	AA400704 zu70h11.r
27	76.6	7.4	740	113	AW202972 fj24g04.y
28	76.4	7.3	307	150	BF562533 UI-R-BUO-
29	76.4	7.3	783	150	BF579861 602095894
30	76	7.3	615	150	BF532437 602074480
31	75.6	7.3	411	141	BE862270 UI-M-BHO-
32	75.2	7.2	658	150	BF531138 602091117
c 33	74.8	7.2	449	11	AA757002 ah54c02.s
34	74	7.1	474	121	AW823937 ufi6e03.y
c 35	74	7.1	537	116	AW450407 UI-H-BF3-
c 36	73	7.0	399	18	AI278269 q178b05.x
c 37	73	7.0	402	11	AA729680 nx37d02.s
c 38	73	7.0	406	16	AI150137 qf44q06.x
c 39	73	7.0	691	169	BF785976 602111621
c 40	72.8	7.0	487	8	AA525010 nh36a02.s
c 41	72.4	7.0	456	17	AI190509 qd49f07.x
42	71.6	6.9	347	7	AA400353 zu64c08.r
43	70.8	6.8	439	7	AA400339 zu64c08.r
44	70.4	6.8	591	137	BE590187 197182 BA
c 45	70	6.7	442	7	AA400181 zu64c08.s

ALIGNMENTS

RESULT 1					
AQ987586					
DEFINITION	AQ987586	566 bp	DNA	30-JAN-2000	
DESCRIPTION	RPCI-23-282C1.TV RPCI-23 Mus musculus genomic clone RPCI-23-282C1, DNA sequence.				
ACCESSION	AQ987586				
VERSION	AQ987586.1	GI:6820791			
KEYWORDS	GSS.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. (bases 1 to 566)				
AUTHORS	Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.				
TITLE	Mouse BAC End Sequences from Library RPCI-23				

JOURNAL COMMENT

Unpublished (1999)  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
7112 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 282 row: C column: 1  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
1. .566  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-282C1"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site:1; EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
BASE COUNT 109 a 137 c 157 g 163 t  
ORIGIN

Query Match	12.9%	Score	134.2	DB	236	Length	566
Best Local Similarity	65.1%	Pred.	No. 5.4e-30				
Matches	256	Conservative	0	Mismatches	113	Indels	24
Gaps	3						
QY	650	aggggtattctgagggccctctgctgtcacattgattggtatgattccagacaggag	709				
Db	81	AGGGGATTCTGGAGGGCCCTCTGCTGTTCACATGTATGCTGTGTAGGCTGATGGTG	140				
QY	710	tagtaagctgggattagatagtggttaaatctcttctctgaggtctacacaaatgaatct	769				
Db	141	TGTTAAAGCTGGGGTTTGGAAATGCGTAAGGATCTTCGGGGTGTGACTCCACGCTGACCT	200				
QY	770	actaccacaaatgagtaataatgccactatttcaagac-----caacaatct	815				
Db	201	ACTACCAAAAGTGGATCAGTGTCTATTATCTCGAGAGCTCTCCGGGGTGGGTGACCA	260				
QY	816	agactctctgactc	875				
Db	261	GCACCCACATGACTCTCTGTTCCCTACTACTATGTTCTCTGCTCTCTCTCTCTCTCTCTC	320				
QY	876	ctgtgctttggaccataacactatacacagatgagctagctgtagctgtagctgtagctg	935				
Db	321	CCGAGGCTTTGGCCCTAGCGCTAATACCAAGTCAATCCA-----GGAGCTGCTGCTGT	372				
QY	936	catacagggctgggaagaagaatgcatgagattagccccagggggagagaactcacagg	995				
Db	373	TGGGGAACTCACCAGGATGGGGTGTGTGGAGAAATAAAGCC---TTAGCGGAGAACTCATGAG	430				
QY	996	agagccactgctaacctcctgggtgactttatta	1028				
Db	431	ACAGTCTCGATGCTGCTGATGACTCCCTTTA	463				
RESULT 2							
AK006271							
LOCUS	AK006271	1050 bp	mRNA	HTC			08-FEB-2001







selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies).

BASE COUNT	88 a	122 c	133 q	129 t
------------	------	-------	-------	-------

Query Match	12.4%	Score 129.4;	DB 238;	Length 472;
Best Local Similarity	68.0%;	Pred. No. 1.5e-28;		
Matches 217;	Conservative	0;	Mismatches 86;	Indels 16;
Gaps 2;				

Qy 650 agggtgattctcgaggccctctgtcggtgcacattgtatgggtgtatggatccagacaggag 709

Qy 710 tagtaagctgggattagaatgtgttaaattcttctcgtggagtctacacccaatgtaattct 769

	DB	TGGTAAGCTGGGTTTGGAAATGCCTGAAGAATTTCCGGGAGATTTCCTCACCAGTAGACCT	217
	Ov	actacccaaaaatggattaatgcccacciatcttcaaggagc-----caacaatatct	815

Db 218 ACTACAAAAGTGGATCAGTCTATTATCTCGAGAGCTCCTCCGGGCTGGGGTGGTGACA 277

Db 278 GCACCCACATGACTTCCTGTCCCTACTATGTACTTTCTCTGGCTCTCCTGGGATCCTT 337

[illegible]

Qy 936 catacaggctgggaagag 954

AL555870	LOCUS	AL555870	876 bp	mrna	EST	16-FEB-2001
DEFINITION	AL555870	LT1 NF1006 PL2	Homo sapiens cDNA clone CS0DK002YM07 5			

ACCESSION	AL55870
VERSION	AL55870.1
GI:	12898012

**SOURCE ORGANISM**

human.  
Homo sapiens  
Chordata: Vertebrata: Euteleostomi:  
Euteleostei: Metazoa:

REFERENCE  
1 (bases 1 to 876)  
Mammalia; Eutheria; Filimales; Catarrhini; Homio.  
Li, W.B.; Gruber, C.; Jesses, J. and Polayes, D.

**JOURNAL COMMENT**  
Unpublished (2001)  
Contact: Genoscope

BP 191 91006 EVRY cedex - France  
Email: [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr).

```

source
1. 8/6
/organism:"Homo sapiens"
/db xref:"taxon:9606"

```

```

/clone_1lib/LTI_NFL006_PL2"
/tissue_type="placenta"

```

was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the NotI and EcoRV sites of the pcwspSP66

Life Technologies. Contact : Feng Liang Life Technologies a division of Invitrogen 9800 Medical Center Drive

Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL : <http://fulllength.invitrogen.com>

BASE COUNT	171 a	251 c	266 g	188 t
ORIGIN				

Query Match	11.2%;	Score 117;	DB 106;	Length 876;
Best Local Similarity	51.8%;	Pred. No. 1.2e-24;		
Matches 394; Conservative	1;	Mismatches 321;	Indels 45;	Gaps

Qy 55 tgtgggcaacctgtatactccagccgcgttgttaggtggccaggatctgtctgcagggcgc 114

Qy 115 tggccttggcaggctcagcctacactttgaccacaactttatctatgagaggttccctcgctc 174

Ov	175	actgaagaggttgatctactgacagcagcacacactgcatacaaacccgacctggagactactttttca	234
DB	164	tggccctgtggcaggaggagccctgcccctctgtgggaattccccacagatgcccggatgagccccatgc	222

Db 224 AGCACCGCTGGGCACCTACGGCGGGGCACCTGCTTTGAAACCTATAGTCACCTTAGTGTAT 283

Db 284 CCCTCCGGTGGATGTTCCAGTTTGGCCAGCTGACTTCCATGCCATCCTCTTGGAGCCTG 343

	-----Tactacgagcccaaaac-----gtcattatatacccccaagtctttaa	329
Qy		289
Dp	344 CAGGCCTACTACACC CGGTACTTCGATCGAATATCTATCTAGCCCTCGGTACCTGGG	403

Qy 328 gatacaacggcagartcgccttgcgaactgtcctctcgaagtcaccttcacttcgcc 387

QY 388 atcctgcctatttgcttgccagtgtcacaaagcagttggcaattccacccttttgg 447

Dd 464 ATCCAGGCCCATCTGTCTCCAGGCCCCTCCACATTTTGASITTAGAACCCGGACAGACATGC 507

Qv 448 qtoacccggatcgggggaaagatttaaggaaaqtttcagataagagaattaccattctgcaccttcag 507

Db 524 GTGACTGGCTGGGGGTACATCAAGAGGATGAGGCACCTGCCATCTCCCCACACCCCTCCAG 583

584 GAAGTTCAGGTGCGCCATCAATAAACAACTCTATGTGCAACCACTCTTCTCAAGTACAGT 643

Qy	568	atattcttgcagcactggagccacgcaaggaaacagaatllctgttcggtcgatc	92
Db	644	TTTCGCAAG-----GACATCTTTGGAGACATGGTTTGTCTGGCAATGCC	688

Qy 628 caaaacatgaaggatagttgcgaagggtgattcttgcgagggcctctgctgctgcacattgat 687

QY 688 ggtgtatggatccagacaggagtagtaagctggggattagaaatgtggtaaatc---tcctt 744

DD 745 cctggagctctacaccaatgtaatactactacacaaaaatggat 785

Db 809 CCGGTGTCTACACCAATATCAGCCACCCTTTGAGTGGAT 849

RESULT 6  
AA266251

**DEFINITION** mz49c09.r1 Barstead mouse pooled organs MPRB4 Mus musculus cDNA clone IMAGE:716752 5' similar to SW:MCP7\_MOUSE Q02844 MAST CELL PROTEIN 7, precursor, mRNA sequence

AA266251.1 GI:1902340

SOURCE	MUSCULUS	ORGANISM
1	<i>Mus musculus</i>	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi

[illegible][illegible]





[illegible][illegible]



Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,  
Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,  
Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,  
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,  
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
Direct Submission  
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9422,  
Fax:81-45-503-9216)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.

FEATURES	Location/Qualifiers
source	1..1256 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /db_xref="MGI:1907454" /db_xref="MGI:1921556" /clone="1700112C13" /sex="male" /tissue_type="testis" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult"
BASE COUNT	287 a 333 c 300 g 316 t
ORIGIN	

Query Match 8.7%; Score 90.2; DB 192; Length 1256;  
Best Local Similarity 47.3%; Pred. No. 2, 5e-16;  
Matches 309; Conservative 1; Mismatches 334; Indels 9; Gaps 1;

Qy	82	gtttagtgccaggatgctgctgcagggcgctggcctggcagggtgcagctacattt	141
Db	173	GTGCTAAATGGGAAGCGGTGGAAGTAGGCAAGTGGCGGTGCAGGTAAGCATTTCTTTC	232
Qy	142	gaccacaacttatctatggaggttccctctcagtgagaggttgatactgacagcacga	201
Db	233	CTGGGAATGTACATCTGCAGCGGTCTCCTCATCCACCACCCTGGATCCTCACCCTGCA	292
Qy	202	cactgcatacaaacgcacctgactacttttcatatctgtgtgctaggatcgattaca	261
Db	293	CAC TGCTTTACAAAGATCCAAACCCGGCTAAATACATCTGTGAAGTGGAGTCCAGACC	352
Qy	262	gtagtgactcaaggaacgtgtgaagtactacgtgtgtccaaaatcgtcatccatcccaag	321
Db	353	CTCCAGACAACAGCACCTCTGAGCTCTGCTGCTACTAGAAATGTAAATTCACGAGAACTTC	412
Qy	322	taccaagatacaacgcgaagcagcagcctgtgtgaaactgtcctctcaagtcacattcact	381
Db	413	ATCAATCCGATGCTGTGACGACATGCCCATCTGAAGCTCAAGATATCTCTACTTGGTCC	472
Qy	392	ctgcccactgcctattgtctggccagtgtcacaaagcagttggccaattccacccctt	441
Db	473	CCCTCTGCACCAATCTGTCTCCCTTCATCAATTTAAAGCCCAAGCATTTGGAACCATG	532
Qy	442	tgttggtgacccgtaggggaaaagttgaagaaagttcagatagagattaccattctgc	501
Db	533	TGCTGGGTCTGGGTGGGACTTCAAAGGCGGAAGGCGACCCCAAGAGATCCCTATAGT	592
Qy	502	cttcgaagcagaagatcaccattattgacgcgcaggtctgtgaacagctctacaatccc	561
Db	593	GTCCAAGGTTTGGTGTTCAGGATTTGTGAACAATGAAAACTGCAATCATCGGTACCA	648
Qy	562	atcggatctctctgcagcagctgagccagtcacaaagcagtaagaaagatttgtctggt	621
Db	649	-----GTTCTCTCTCTGCTGAAGAACACAGAAAAGTTTCATTTGGGAACGACATGTTGTGTACA	703
Qy	622	gatactcaaacataaaggatagttgtcaagggtgattctggaggggcctctgtcgtgtcac	681

Db 704 AGCTAGAACTGGCGCTGGACACTGTTCAGGACACCTCCGGAACTCTCTGGTTGCCAG 763

Qy 682 attgatggtgatgataccagacaggagtagtaagctggggatcagtagatg 734

Db 764 ATGAACAAGACCTGGTCCAGATGGCGCTGGTGGAGCTTGTGACTGTGG 816

RESULT 14

AK010640 1629 bp mRNA HTC 08-FEB-2001

LOCUS Mus musculus ES cells cDNA, RIKEN full-length enriched library,  
clone:2410039E18, full insert sequence.

DEFINITION AK010640

ACCESSION AK010640.1 GI:12846228

VERSION CAP trapper.

KEYWORDS Mus musculus (strain:C57BL/6J) ES cells cDNA to mRNA,  
clone\_lib:RIKEN full-length enriched mouse cDNA library

SOURCE clone:2410039E18.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (sites)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning  
Methods Enzymol. 303, 19-44 (1999)

2 (sites)  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

20499374

3 (sites)  
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,  
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,  
Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,  
Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,  
Kashiwagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,  
Matsura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

20530913

4 (sites)  
The RIKEN Genome Exploration Research Group Phase II Team and  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

5 (bases 1 to 1629)  
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,  
Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,  
Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Horii,F.,  
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,  
Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,  
Miyaizaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,  
Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,  
Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Suzuki,H., Tagami,N., Tagawa,A., Takahashi,F.,  
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,  
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama  
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)

PLEASE VISIT OUR WEB SITE (<http://genome.gsc.riken.go.jp/>) FOR  
further details.

COMMENT cDNA library was prepared and sequenced in Mouse Genome



Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5', GAGACGAGAGAGGATCAAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGACGAGAGATTCGAGTTAAATTAAATCCGCCGCCGCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

```

FEATURES
source
Location/Qualifiers
1. .1629
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:GRI:1910128"
/db_xref="MGD:GRI:1923810"
/clone="2410039E18"
/cell_type="ES cells"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
BASE COUNT
317 a 501 c 433 g 378 t

```

Query Match	8.6%	Score 89.4;	DB 192;	Length 1629;
Best Local Similarity	50.4%	Pred. No. 4.8e-16;		
Matches 329; Conservative	1;	Mismatches 307;	Indels 16;	Gaps 4;

QY	79	cgcgctgtaggtgcccaggatgctgctgcgagggcgctggtcgttcgaggtgcagcgaactaac	138
Db	244	CGATCATCCGGTGGTGCGATCGAAAGCCGGTCAGTGGCCCTGGCAGTCAGCATCACC	303
QY	139	tttgaccacaaacttactatgagaggttcctcgtcgtcagtgagaggttgatactacagca	198
Db	304	TACGATGCCAACCATGTTTGTGGGGGTGCGCTCGTGTCAATAATGGTGGTGTCTGTGCT	363
QY	199	gcacactgcatacaacgcgacctggactacttttctatatactgtgtgtaggatacgatt	258
Db	364	GCTCACTGCTTCCCCAGAGAACACAGACAGGGAAGCGGTATGAGGTGAAGCTGGGGGCCAC	423
QY	259	acagtaggtgactcaaggaacg---tgtagagtaactacgtgtccaaatcgtcatccat	315
Db	424	CAGCTAGACTCCFACAGCAATGACACTGTGGTCCACACAGTGGCTCAGATCATCACCCAC	483
QY	316	ccaagtaccagaata-----caacggcgaacrtgccttgttgaaactgtcctctcaa	369
Db	484	TCAAGCTTACCAGGAAGAGGGCTCCACAGGGGACATCGCGTCTATCCGCCCTCAGCAGTCT	543
QY	370	gtcaccttcaactctgcattcctctgcttattgtctgccagtggtcacaaagcagattggca	429
Db	544	GTACACTTCTCCGGCTACATCAGACCCTATGCTCCTCCTGCAGCCAAATGCTCCTTTCC	603
QY	430	attccaccttbtgttggtgcacggatgggggaaagttaaggaaagtccagatagagat	489
Db	604	AACGGCCTTCACTGTACTGTACGGGATGGGGTCATGTGGCTCCTTCAGTGAGGCTCCAG	663
QY	490	taccattctcccttcagggaagcagaagaatcaccattattgacgcgcaggttgttgaacag	549
Db	564	ACCCCTAGGCCCTTGCAGCAGCTCAGGTACCACTCATCAGCCGGGAACCTGTAGGTGC	723
QY	550	ctctacaatcccactcggtatcttcttccagcactgagccagtcatacgaaggaagaag	609
Db	724	CTGTACAA-----CATTAATGGGTGCCCTGAAGAACCGCACACTATCCAGCAGGCATG	777
QY	610	atttgctgtagtaactcaaacatagaagatagttgcgaaggtgattcttgaagggcct	669
Db	778	CTGTGTCTGGCTAT-GTGAAGGGAGGACAGATGCCCTGCCAGGGTGACTCTGGGGGGCCCA	836
QY	670	ctctgctgtcacattgatggttatggattccagacagagtagtaagctgggg	722

Db 837 CTCCTTGTCCTCCCATGGAGGGCATCTGGTACTTGGCAGGCATTGTGAGTTGGGG 889

RESULT 15  
AI326140/C

LOCUS	EST	23-DEC-1998
AI326140	542 bp	mRNA
mr64g08.x1	Stratagene mouse testis (#937308)	Mus musculus CDNA
clone IMAGE:602270	3' similar to SW:PSS8_HUMAN Q16651	PROTASTIN
	PRECURSOR ;	mRNA sequence.

ACCESSION AT326140  
VERSION AT326140.1 GI:4060569  
KEYWORDS EST.  
SOURCE house mouse.

ORGANISM

100

REFERENCE  
AUTHORS

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)

CONTACT: Marra M/Mouse EST Project  
WashU-HIMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Fax: 314 200 1010  
 Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:367702

This clone was previously sequenced on the 5' end only, this new data is from the 3' end  
Possible reversed clone: similarity on wrong strand  
High quality sequence stop: 444.

FEATURES SOURCE

```

/organism="Mus musculus"
/strain="Inbred CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:602270"
/clone_lib="Stratagene mouse testis (#937308)"
/sex="males"

```

```
/tissue_type="testis"  
/day_age="10-12 week old"  
/lab_pos="SOLR kanamycin resistant"  
/note="Organ: testis; Vector: pBlueScript SK+;  
ECORI_Site_2_XhoI_Cloned undifferentiated  
Oligo dn Average insert size: 1.0 kb; UniAP  
-5' adaptor sequence: 5' GAATTCGGCAGG 3'  
sequence: 5' CTCGCACGTCTTCCATTCACTCGCATTTTG
```

BASE COUNT  
ORIGIN

Query Match	8.6%	Score 89.2;	DB 18;	Length 542;
Best Local Similarity	55.1%	pred. No. 3.8e-16;		
Matches 228:	Conservative			
	0: Mismatches 168;			Indels 18;

Qy 387 catcctgcctatttgcttgcgccagtgccacaagcagttggcaattccacccttttgtg 446

Db 540 CATCCAGCCCATCTGCTCTCTGAACTCCACGTAACAAGTTTGAGAACCGAACTGACTGCTG 481

Qy 447 ggtgaccgcatgggggaaaagttaaggaaagttcagatagagattacattctgccttca 506

Db 480 GGTCACGGCTGGGGGGCTATTGGAGAAGATGAGAGCTGCCCATCTCCCAAGACTCTCCA 421

Qy 507 ggaagcagaagtaccattattgaccgccaggcttgtgaacagctctacaatcccatagg 566

Db 420 GGAAGTGCAGGTAGCTATTATCAACAACAGCATGCTAACCATATGTACAAAA----- 365

Mon Nov 26 09:37:53 2001

QY 567 tatcttcttccagcactggagccagtcacaaagacaagatttgtgctgggtgatac 626  
Db 366 -----GCCAGACTTCCGACGACATCTGGGAGACATGGTTTGGGCTGGCACTCC 316  
QY 627 tcaaaacatgaaggatagttgcaagggtgattctctggaggccctctgtctgcacattga 686  
Db 315 TGAAGGTGGCAAGGATGCCTGCTTTGGTGACTCGGGAGGACCCTTGGCCTGGGACCAGGA 256  
QY 687 tgggtgtatggtatccagacagagtagtaagctggggattagaaatgtggtaaatc---tct 743  
Db 255 TAGCGTGTGTATCAGGTTGGAGTTGTGAGCTGGGGAATAGGCTGTGGTCGCCCAATCG 196  
QY 744 tcctggagtctcacaccaatgtaattctactaccacaaatggatttaatggccactat 797  
Db 195 CCCTGGAGTCTATACCAACATCAGTCATCATTACCAACTGGATCCAGTCAACCAT 142

Search completed: November 21, 2001, 23:54:06  
Job time: 1401 sec

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2001, 23:34:11 ; Search time 127.85 Seconds  
(without alignments)  
5112.597 Million cell updates/sec

Title: US-09-854-844-1  
Perfect score: 1041  
Sequence: 1 atgggcctctgctgtgc.....ttattacaatttgaaatga 1041

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	132.8	12.8	942	22	AAF75647 Murine ztrypl dege
2	129.4	12.4	959	19	AAV59132 Nucleotide sequenc
3	118.6	11.4	942	20	AAV59136 Nucleic acid encod
4	118.6	11.4	1082	20	AAV59133 Nucleic acid encod
5	118.6	11.4	1100	19	AAV59119 Nucleotide sequenc
6	117.2	11.3	1079	21	AAZ45673 Nucleotide sequenc
7	117	11.2	942	21	AAV62005 Hydrophobic domain
8	117	11.2	1081	20	AAV87151 Human cancer-speci
9	117	11.2	1081	21	AAA08505 Hydrophobic domain
10	117	11.2	1087	21	AAA62015 Nucleotide sequenc
11	117	11.2	1117	19	AAV59118

pot ease activity  
by homology

12	115.8	11.1	1110	22	AAF76994 Human protease T c
13	115.8	11.1	1703	21	AAZ52473 HTRM clone 3376404
14	115.4	11.1	1100	20	AAV52259 Protein PRO303 CDN
15	115.4	11.1	1100	22	AAF72417 Human PRO303 CDNA.
16	114	11.0	980	19	AAV59136 Nucleotide sequenc
17	112.6	10.8	1103	22	AAF55271 Nucleotide sequenc
18	110	10.6	1130	22	AAF77000 Fusion gene of pro
19	98.6	9.5	936	22	AAF75659 Human ztrypl degen
20	96.2	9.2	1154	22	AAF75646 Murine ztrypl codi
21	95	9.1	1142	21	AAV87796 Activation constru
22	95	9.1	1142	22	AAF55268 Nucleotide sequenc
23	95	9.1	1169	21	AAV87795 Activation constru
24	95	9.1	1169	22	AAF55267 Nucleotide sequenc
25	93	8.9	1668	19	AAV77814 Human cancer assoc
26	91.2	8.8	1097	19	AAV44325 Rat homologue of m
27	91.2	8.8	1108	19	AAV44333 Murine mMCP-6 zymo
28	91.2	8.8	1108	19	AAV44333 Mouse mast cell pr
29	90.4	8.7	933	19	AAV59135 Nucleotide sequenc
30	89.6	8.6	1103	19	AAV42714 Rat homologue of m
31	81.8	7.9	1031	19	AAV44323 Murine mMCP-7 zymo
32	81.2	7.8	1322	21	AAV61704 cDNA encoding mous
33	76.4	7.3	1221	20	AAV41376 Human normal uteru
34	73	7.0	1305	22	AAV02556 Human seripancrin
35	73	7.0	1479	22	AAV02557 Human seripancrin
36	73	7.0	2038	20	AAV87154 Human protease HUP
37	73	7.0	2063	21	AAV37099 Human PRO1570 (UNQ
38	73	7.0	2063	22	AAV92113 Human PRO1570 CDNA
39	73	7.0	2063	22	AAV54396 DNA encoding prote
40	72.8	7.0	1137	19	AAV44329 Human mast cell tr
41	72.8	7.0	1137	19	AAV42711 Human mast cell tr
42	72.8	7.0	1137	21	AAV21079 Human low adenosin
43	72.8	7.0	1137	21	AAV34957 Human adenosine re
44	72.8	7.0	1154	19	AAV44328 Human mast cell tr
45	72.8	7.0	1154	19	AAV42710 Human mast cell tr

ALIGNMENTS

RESULT 1  
ID AAF75647 standard; DNA; 942 BP.  
XX AAF75647;  
XX AAF75647;  
DT 10-MAY-2001 (first entry)  
XX Murine ztrypl degenerate coding sequence.

XX Mouse; ztrypl; serine protease; tryptase; inflammation; fertilisation;  
KW cardiovascular disease; infertility; asthma; immune disorder; stroke;  
KW gastrointestinal disorder; testicular function; contraceptive; ds.  
XX Mus musculus.

OS Mus musculus.  
PN WO200112788-A2.  
XX 22-FEB-2001.  
XX 09-AUG-2000; 2000WO-US22156.  
XX 18-AUG-1999; 99US-0376445.  
XX (ZYMO ) ZYMOGENETICS INC.  
XX Presnell SR, Taft DW;  
XX WPI: 2001-202859/20.  
XX P-PSDB; AAV72885.  
XX New mouse serine protease polypeptides ztrypl and polynucleotides,  
PT useful for treating cardiovascular disease, infertility, impotence and  
PT other male reproductive dysfunction -

XX Disclosure; Page 102; 112pp; English.

XX PS

XX CC The present invention provides the protein and coding sequences of the

CC human and murine serine protease ztryp1. This is a trypsin like protein

CC which is highly expressed in contractile tissues. The sequences can be

CC used in the treatment and identification of treatments for cardiovascular

CC disease, inflammation, infertility, male reproductive dysfunction, in

CC asthma, stroke, immune disorders and gastrointestinal disorders. In

CC addition, they can be used to modulate testicular function and as

CC contraceptives.

XX SQ Sequence 942 BP; 160 A; 87 C; 153 G; 144 T; 398 other;

Query Match 12.8%; Score 132.8; DB 22; Length 942;

Best Local Similarity 33.7%; Pred. No. 1.5e-31;

Matches 253; Conservative 122; Mismatches 362; Indels 12; Gaps 3;

QY 44 ggatctcagtggtggcacaactgtatctaccagccgctgtgtggtggccaggatgctg 103

DB 92 ggathtgwsntgygncarcacnaaya thacntgyaargtngtnaayggnaargcngtng 151

QY 104 ctgcaggcgctggccttggtggcaggtcagctacatcttgaccacacatttactatggag 163

DB 152 argtngngaartggcctggcargctwnsnathynttvtngnngnatgaya chtgwsng 211

QY 164 gttccctcgctgaggggtgtatctacgcagcactgcacatctgcatacaaccgacctga 223

DB 212 gnwsyntnathcaycaycaytggathytnacngcngcncaytgyytnearmgnwnaara 271

QY 224 ctacttttctatctgtgtggtcaggtgattacagtgaggtgactcaagaagaaactgctg 283

DB 272 aycngcnaartayacngtnaargtngngtncaracnvtncngngayaaywsnacnwsng 331

QY 284 tgaagctactaggtcccaaaatcgctacatccatcccaagatacaagaacagcgacagac 343

DB 332 ariytnyngtnacmngnathtgna thcaygaraytytthaaymgnatgwsngaygaya 391

QY 344 tgccttgttgaaactgtcctctcaagtcacactctcactctgccatccctgcctatttgcct 403

DB 392 thgcnathytnaarytnaartaycngtnaactggwscnctyngtngtncarcnathgtgy 451

QY 404 tggccagtgctacaaagcagttggcattccaccctctttgttgggtgcagcgatggggaa 463

DB 452 tncnwsnttyaayytnaarccnwnathtgnaacnattgtygggtngtnggtngggngny 511

QY 464 aagttaaaggaagtccagatagattaccattctgccttcaggagcagagaagtaccaca 523

DB 512 tngaraargcngargncayccnaaracnctaywsngtncarggnytnngcngtnmna 571

QY 524 ttatgacgcgcagctgttgagacagctctacaa tcccatcggtatctcttctgccagcac 583

DB 572 thgtnaayaagarathgtayaaycaymgntayca -----rttyytnytnaaraayc 625

QY 584 tggagccagtcataaggaagacaagattgtgctggtgatactcaaacatgaagagata 643

DB 626 aaraaarttyathgnaaygayatgtytngtyacnws ---nwsngartgggngtyngaya 682

QY 644 gttcaaggggtattctgagagggcctctctggtgtcacattgattggtatgatgatccaca 703

DB 683 cntgycargayacnwsnggnwnswntgntngtgcaratgaaayaaacnctgggtncara 742

QY 704 caggagtagtaagctgggattagaaatgggttaaatctc ---tctctggagtgatcacacca 760

DB 743 tgggngtngtnwsntgaaaytyggytgytgggngmngmncartcyccnwsngtntayaacnw 802

QY 761 atgtaatctactaccacaaatggattaca 788

DB 803 snacnwsncaytyacncartggtathaa 830

## RESULT 2

```
Db 103 ttatggtgggaaccacttatgtggcgcaacctgtgtcaaccgcgctgggtgtctacagc 162
Qy 198 agcacactgcatacaacgcagcctgactacttttcaatactgtgtgctaggtacgat 257
  ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 163 tcccaactgtcccaaaaggaataacgactctttgtactggacagctccagttgtgagct 222
Qy 258 tacagttagtgactcaaggaacgtgtgaagtacta-----cgtgtccaaaatcgtca 310
  || || || || || || || || || || || || || || || || || || || || ||
Db 223 gacttccaggccatctctggaacctacagcgctattccaaccgttaccaaaatagaaga 282
  || || || || || || || || || || || || || || || || || || || || ||
Qy 311 tcatcccaagtaccagaatacaaacggca-----gaactgccttgtgaaact 359
  || || || || || || || || || || || || || || || || || || || || ||
Db 283 tatttctgagcccaagactctcggagcagtagtaccatgacatgacactgctgaagct 342
Qy 360 gtccctcaagtcactcaactctgcctcctgcctatttctgttgcctgctccaggtcacaaa 419
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 343 gtcactccagtcactcaataaactctaccagccatctgcctcctcctgaaactccacgta 402
Qy 420 gcagttggcaattccaccctttgttgggtgaccggatggggaagtttaaggaagttc 479
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 403 caagtttgagaaccgaactgactgctgggtgaccgctgggggctatttgagaagatga 462
Qy 480 agatagagattaccattctgccttcaggaagacagaagtaaccattattgaccgcagggc 539
  || || || || || || || || || || || || || || || || || || || || ||
Db 463 gagtctgcatctcccaacactctccaggaagtcgaggtagctattatcaacaacagcat 522
Qy 540 ttgtgaacagctctacaatcccatcggtatcttctgcagcactggagcagtcatacaa 599
  || || || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 523 ggttaaccatagtcaaa-----aaagccagacttccgcacgaacatctg 567
Qy 600 ggaagacaagatttgcgtggtgatactcaaaacatgaaggatgttgcgaaggtgattc 659
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 568 gggagacatggttgcgtgcactccctgaaggtggaagtgctgttggtagctc 627
Qy 660 tggagggcctctgctgcattgatgtgtatggatccagacagagtagtagtg 719
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 628 gggagggaccctggctgcagcaggatcaggtgtggtatcaggttgaggtgtgagctg 687
Qy 720 gggattagaatgtgaaatc---tctcctggagtcacacaaatgtaactactacca 776
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 688 gggaaataggtgtgtgcgccaaatgcctggagtcttataccacaatcagtcactca 747
Qy 777 aaaaatggattaatgcactat 797
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 748 caactggatccagtcacacat 768

RESULT 3
AAAX15336
ID AAAX15336 standard; cDNA to mRNA; 942 BP.
XX AC AAAX15336;
XX AC AAAX15336;
XX DT 04-MAY-1999 (first entry)
XX DE Nucleic acid encoding a human eosinophil serine protease.
XX KW Human; eosinophil; serine protease; allergic disease; infectious disease;
XX KW tumour; granulomatous disease; collagen disease; vascular inflammation;
XX KW ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 1..942
XX FT /*tag= a
XX FT /note= "no termination codon"
XX PN JP11032768-A.
XX PD 09-FEB-1999.
XX PF 16-JUL-1997; 97JJP-0191319.
```

```
XX 16-JUL-1997; 97JJP-0191319.
PR (ONOY ) ONO PHARM CO LTD.
PA WPI; 1999-183825/16.
XX P-PSDB; AAW97116.
DR New eosinophils serine protease - useful for prevention and
PT treatment of allergic, infectious, tumour, granulomatous and collagen
PT diseases
XX Claim 4; Page 10; 18pp; Japanese.
XX The present sequence encodes a human eosinophil serine protease.
CC The protease is useful in drug compositions for the prevention and
CC treatment of allergic diseases, infectious diseases, tumour diseases,
CC granulomatous diseases, collagen diseases and vascular inflammation.
XX Sequence 942 BP; 176 A; 283 C; 276 G; 207 T; 0 other;
SQ

Query Match 11.4%; Score 118.6; DB 20; Length 942;
Best Local Similarity 51.9%; Pred. NO. 4.3e-27;
Matches 395; Conservative 1; Mismatches 320; Indels 45; Gaps 4;

Qy 55 ttgtggcaacctgtatactccagccgctgttagtggccagagatgctgctcagggcgc 114
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 97 tgcggccgacgggtcatcactgcgcgcatcgtgggtggagggagcgcgaactcgggcgt 156
Qy 115 tggccttggcaggtcagcctacactttgaccacaactttatctatggaggttccctcgtc 174
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 157 tggcgtggcagggagcctgcctgtgggtatccacgtagtcggagtgagcctgctc 216
Qy 175 agtgagaggttgcactgacagcagcacaactcacaacacgacctggactacttttca 234
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 217 agccaccgtggcactcagcgcgctgttgaacctatagtgacctagtgt 276
Qy 235 tatactgtgtgctaggtacgtatgagtgagtcacaggaagaaacgtgtgaag----- 288
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 277 cctcgggtggatggtccaggttggcagctgacttccatgccactcttctggagcctg 336
Qy 289 -----tactacgtgtccaaaatc---gtcatccatcccaagtaccaa 327
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 337 caggcctactacacccgttactctgtatcgaatatctatctatgacctcgtacctgggg 396
Qy 328 gatacaacggcgaortgccttgttgaacctgtcctctcaagtcacacttcaactctgccc 387
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 397 aattcacctatgacattgccttgggaagctgtctgcacctgtcacctacacataaac 456
Qy 388 atcctgcctatttgcctccaggtgtcacaaagcagttggcaatccacccttttgg 447
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 457 atccagcccatctcctccaggcctccatcttggagttgagacggacagactgtctgg 516
Qy 448 gtgaccgtaggggaaaagttaaggaaagtctcagatagagattaccattctgcttccag 507
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 517 gtgactggtgggtatcatcaaaagagtaggagcactgccatctcccccacacctccag 576
Qy 508 gaagcagaagtaccattattgacgcgaggttgtgaacagctctacaatccccctgg 567
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 577 gaagtccaggtgcctatcaataaacactctatgtgcaaccctcttctcgaagtacgt 636
Qy 568 atctcttgcagcactggagccagtcacaggaagacaagattgtgtggtgatact 627
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 637 ttccgcaag-----gacatcttggagacatggttgtgtgctggcaatgcc 681
Qy 628 caaacatgaagtagtggcaagggtgattcttgagggcctctctgctgtacacattgat 687
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 682 caaggcgggaagtagccttgcctgactcagtgaggaccttggcctgtgaacaagaat 741
Qy 688 ggtgtatggtaccagacagagtagtaagctgggagattagaatgtggtaaatc---tctt 744
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 742 ggactgtggtatcagattgagtcgtgagctggggagtggtgtggtgcgccaatcgg 801
```



XX  
PS Claim 6; Pages 62-64; 167pp; English.

The present sequence represents the nucleotide sequence of the long isoform of HELA2. cDNA generated from HeLa cells expressing HELA2 cells was amplified using PCR primers AAV48312-13. Three new sequences were detected in the 480 bp amplicon. These sequences are designated HELA2 and ATC2 which have high homology to serine proteases and BCOM3 which has homology to a kinase. The proteins are involved in or associated with regulation of cell activity and/or viability. Administration of recombinant HELA2 (also called testisin) is used to increase fertility. Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of testicular germ cell cancers (seminoma) and is also expressed in some non-testicular cancers (of colon, pancreas, prostate and ovary), so is a marker/potential therapeutic target for cancer. The promoter from the HELA2 gene is useful for testis-specific expression of other genes, e.g. for gene therapy or modulation of fertility. Drugs that block activity of HELA2 should have antitumour activity (other than in testis) while in testis recombinant HELA2 should stop growth of tumours and normalise sperm development (eliminating the need for orchidectomy). Identification of mutant forms of HELA2 can be used to diagnose infertility.

Sequence 1100 BP; 226 A; 319 C; 313 G; 242 T; 0 other;

Query Match	11.4%	Score 118.6;	DB 19;	Length 1100;
Best Local Similarity	51.9%	Pred. No. 4.7e-27;		
Matches 395;	Conservative	1;	Mismatches 320;	Indels 45;
				Gaps 4

Qy 55 t g t g g g c a a c c t g t a t a c t c c a g c c g c g t t g t a g g t g g c c a g g a t g c t g c t g c a g g g c g c 114

[illegible]

0y 115 tggccttggcaggtcagccctacaccttgaaccacaaacttatctatggaggttccctcgtc 17A

[illegible]

\_\_\_\_\_

QY I/5 agtgagaggttgatactgacagcagcacactgcatacaaccgacctggactacttttca 234

233 ayyccaccyctcyggcacctccacggyccygcacccctacatagctgat 292

Qy 235 tatactgtgtggctaggatcgattacagtaggtgactcaaggaaacgtgtgaag----- 288

Db 293 ccctccggatgagtccagttttagccagctgacttccatgccatcccttctgagacctg 352

[illegible]

UY 289 -----tactacgtgtccaaatc---gtcatccatcccaagtaccaa 327

Db 353 caggcctactaccccgttacttcgtatcgaatatctatctgagccctcgctacctggg 412

Oy 328 gatacaacggcagacrtcgcccttgttgaaactgtcctctcaagtcaccttcacttctgcc 387

— — — — —

Db 413 aattcacccctatgacattgccttggtgaagctgtctgcacctgtcacctacactaacac 472

QY 388 atcctgcctatttgcttgcccagtggtcacaagcagtggtggcaattccacccttttgttgg 447

Db 473 atccagcccatctgtctccaggcctccacatttgagtttgagaaaccggacagactgctgg 532

[illegible]

Qy 448 gtgaccggatgggggaaagtttaagggaagtttcagatagagattaccattctgccccttcag 507

Db 533 gtgactggctgggggtacatcaagaggatgaggcactgccatctccacacccctccag 592

Oy 508 gaagcagaagtacccattattgaacggccaggcttataaacaggtctacaatcccatcagt 567

[illegible]

```

Db 844 AGCCACCGCTGGGCACTACACGGCGCGCACTGCTTTTGAACCTATAGTGACCTTAGTGAT 785
Qy 235 tatactgtgtgctaggatcgtattacagtagtgactcaagaaacgtgtgaag----- 288
Db 784 CCCTCGGGTGGATGGTCAGTTTGGCCAGCTGAGACTTCCATGCCATCCCTTCGTGGAGCCTG 725
Qy 289 -----tactacgtgtccaaaaatcg-----tcattccatcccaagtaccaa 327
Db 724 CAGGCTACTACACCGTTACTTCNTATCGAATATCNATCTGAGCCCTCGCTAACTGGGG 665
Qy 328 gatacaacggcagactcgcctctgttgaacatgtctctcaagttcaactctcaactcttgc 387
Db 664 AATTCAACCTATGACATTCGCTTGGTGAAGCTGCTGCAACCTTCACCTTACACTAAACAC 605
Qy 388 atcctgcctattgtctgcccagtgatcaaaagcagttggcaattccacccttttgg 447
Db 604 ATCCAGCCCATCTGCTCCAGGCTCCACATTTGAGTTTGAACCGGACAGCTGCTGG 545
Qy 448 gtaccggtatggggaaagttaaggaagtccagatagagattaccattctgccttcag 507
Db 544 GTGACTGGCTGGGGTACATCAAAAGAGGATGAGGCACCTGCCATCTCCCCACACCCCTCCAG 485
Qy 508 gaagcagaagtaccattattgaagcgcggcttgtgaacagctctacaatccccatcggt 567
Db 484 GAAAGTTCAGGTCGCCATCATAAACAACTCTATGTGCAACCCACTCTTCCTCAAGTACAGT 425
Qy 568 attctttgccagactggagccagtcacatcaaggaagacaagaattgtgtgtgatact 627
Db 424 TTCGGCAAG-----GACATCTTTGGAGACATGTTTGTCTGCGCAATGCC 380
Qy 628 caaaacatgaagatgttgcaagggtgattctgggggctctgtcgtgtgcacattgat 687
Db 379 CAAGCGGGAAGGATGCTGCTTCGGTGAAGTCAAGTGGACCCCTTGCCTGTAAACAAGAT 320
Qy 688 ggtgtatggtatccagacagagtagtaagctggggattagaattggtgtaaatc---tcctt 744
Db 319 GGACTGTGTGATCAGATTGAGATTCGTGAGCTGGGGAGTGGGCTGTGGTGGNCCAATCGG 260
Qy 745 cctggaggtacaccaagttaactactaccacaaaattggat 785
Db 259 CCCGGTGTCTACACCAATATCAGCCACCACTTTCAGTGGAT 219

RESULT 7
AAA62005
ID AAA62005 standard; DNA; 942 BP.
XX
XX AAA62005;
AC
XX
XX
XX 02-FEB-2001 (first entry)
XX
XX Hydrophobic domain protein cDNA HP03116 isolated from KB cells.
XX
XX Human; secreted protein; membrane protein; hydrophobic domain;
KW proliferation control; differentiation induction; material transport;
KW biophylaxis; signal receptor; ion channel; transporter; immunostimulant;
KW immunosuppressant; haematopoiesis regulator; chemotactic; chemokinetic;
KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;
KW autoimmune disease; Alzheimer's disease; Parkinson's disease; cancer;
KW gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO200029448-A2.
XX
XX
XX 25-MAY-2000.
XX
XX 17-NOV-1999; 99WO-JP06412.
XX
XX 17-NOV-1998; 98JP-0326255.
XX
XX 22-DEC-1998; 98JP-0364315.
XX
XX 16-MAR-1999; 99JP-0069811.
XX
XX 27-APR-1999; 99JP-0119299.
XX
```

```

PR 19-MAY-1999; 99JP-0138169.
XX
XX (SAGA ) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
XX
XX Kato S, Kimura T;
PI
XX WPI; 2000-387753/33.
DR P-PSDB; AAB12132.
XX
XX Proteins comprising hydrophobic regions, such as secretory and membrane
PT proteins, useful in research and diagnostics and having various
PT activities e.g. immunomodulatory, antiinflammatory, chemokinetic,
PT hemostatic, thrombolytic -
XX
XX Claim 3; Page 250; 410pp; English.
XX
XX Secretory proteins play important roles in the proliferation control, the
CC differentiation induction, the material transport and the biophylaxis of
CC cells. Membrane proteins have important roles as signal receptors, ion
CC channels and transporters. The present sequence is the coding sequence
CC for a human protein which has at least one hydrophobic domain. The
CC protein encoded by the present sequence may be a secretory or a membrane
CC protein. The encoded protein may have cytokine and cell
CC proliferation/differentiation activity, immune stimulating or suppressing
CC activity, haematopoiesis activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, anti-inflammatory activity and tumour
CC inhibition activity. The present sequence could therefore be used for
CC treatment of autoimmune disease, Alzheimer's disease, Parkinson's
CC disease, and cancer via gene therapy.
XX
XX Sequence 942 BP; 175 A; 282 C; 277 G; 208 T; 0 other;
SQ

Query Match 11.2%; Score 117; DB 21; Length 942;
Best Local Similarity 51.8%; Pred. No. 1.4e-26;
Matches 394; Conservative 1; Mismatches 321; Indels 45; Gaps 4;

Qy 55 tgtgggcaacctgtatactccagccgctgttagtggtggcaggatgctgtcgagggcgc 114
Db 97 tgcggcgacgggtcatcactcgctcgcatcgtgtgggtggagaggacgccgaactcgggcgt 156
Qy 115 tggccttggcaggtcagctacactttgaccacaaactttatctatgagggttccctgc 174
Db 157 tggccgtggcagggagcctgcgctgtgggtattccacagtatgcggagtgcctgc 216
Qy 175 agtgagaggttgatagtcagcagcagcactgcatacaaccgacctggaactatttca 234
Db 217 agccaccgtgggcactcacggcgccgctgtttgaaacctatagtgacctagtgtat 276
Qy 235 tatactgtgtggctaggatcgattacagtaggtgactcaaggaaacgtgtgaag----- 288
Db 277 cctccgggtggatgggtccagtttggcagctgacttccatgccatccttctggagcctg 336
Qy 289 -----tactaogtgcacaaatc---gtcatccatcccaagtaccaa 327
Db 337 caggcctactacacccgttactctgtatcgaatactatctgtgagccctcgtacctggg 396
Qy 328 gatacaacggcagactcgtgttgaacactgcctcctcaagtcaacttcaacttgc 387
Db 397 aattcacccctatgacattgcttgggtgaagctgctgcacctgtccactacactaacac 456
Qy 388 atcctgcctatttctgtgccagtgtcacaaagcagttgccaattccacccttttgg 447
Db 457 atccagccctctgtctccaggtccacattgtgagtttgagaaccggacagactgctgg 516
Qy 448 gtgaccggatggggaaagttaaggaagtccagatagagattaccatttgccttcag 507
Db 517 gtgactggctgggggtacatcaaaagagatgagcactgccaactctccccacacctccag 576
Qy 508 gaagcagaagtacccattattgaccgcaggctgtgaaacagctctacaatccccatcggt 567
XX
XX
```



Db 577 gaagttcagtcgcatataaacaacttatgtgcaaccactcttctcactgaagtagt 636  
 Qy 568 atcttctccagcactgagccagtcatacaaggaagacaagaattgtctgtgtact 627  
 Db 637 ttccgcaag-----gacactttggagacatggtttgtctgtggaatgcc 681  
 Qy 628 caaacatgaagtagtgcaagggtgattcttgaggccctctctgtctgtcacattgat 687  
 Db 682 caaggcggaagtagctcttcgtgtagctcaggtggacccttgccctgtaacaagaat 741  
 Qy 688 ggtgtatgatccagacagtagtagtaagtagctgggattagaattgtgttaaatc---tctt 744  
 Db 742 ggaactgtgtatcagatgagtcgtgagctgggagtggtgtggtcgcccaatcgg 801  
 Qy 745 cctggagctacaccaatgaatctactaccacaaatgat 785  
 Db 802 cccggtgtctacaccaatcagccaccactttgagtggaat 842

## RESULT 8

AAx87151  
 ID AAX87151 standard; cDNA; 1081 BP.

AC AAX87151;

DT 27-SEP-1999 (first entry)

XX Human protease HUPM-3 cDNA.

XX Protease; human; HUPM-3; cell proliferation; cancer;

KW immune disorder; inflammation; therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 40..984

FT /\*tag= a

FT sig\_peptide 40..76

FT /\*tag= b

FT /\*note= "putative signal peptide sequence"

FT mat\_peptide 77..981

FT /\*tag= b

XX W09936550-A2.

PN 22-JUL-1999.

XX 12-JAN-1999; 99WO-US00655.

XX 16-JAN-1998; 98US-0008271.

XX (INCY-) INCYTE PHARM INC.

PA Bandman O, Corley NC, Guegler KJ, Hillman JL, Shah P;

PI Tang YT, Yue H;

XX WPI: 1999-430616/36.

DR P-PSDB; AAY06434.

XX Novel human protease molecules useful in the treatment of

PT developmental disorders and/or cancers

XX Claim 8; Page 84-85; 90pp; English.

XX This nucleotide sequence codes for HUPM-3 (see AAY06434), a novel

CC human protease. HUPM-3 cDNA was initially identified in Incyte

CC Clone 789927 from the prostate tumour cDNA library PROSTUT03 using

CC a computer search for amino acid sequence alignments. The

CC present sequence is a consensus sequence derived from the following

CC overlapping and/or extended nucleic acid sequences: Incyte Clones

CC 789927, 1646976 (PROSTUT09) and 1979791 (LUNGUT03). A fragment

CC comprising nucleotides 271-330 of the present sequence can be used

CC for hybridisation. Northern analysis shows expression of this

CC sequence in cardiovascular, haematopoietic and male reproductive  
 CC cDNA libraries. Approximately 86% of these libraries are  
 CC associated with neoplastic disorders. The invention provides 12  
 CC new human proteases, i.e. HUPM-1 to -12 (see AAY06432-43), and the  
 CC polynucleotides encoding them (see AAX87149-60). Also provided are  
 CC vectors, host cells and methods for producing HUPM polypeptides,  
 CC as well as agonists and antagonists of HUPM. Methods for treating  
 CC or preventing cell proliferative disorders and immune disorders  
 CC using HUPM or HUPM antagonists are claimed.

XX  
 SQ Sequence 1081 BP; 202 A; 321 C; 321 G; 237 T; 0 other;

Query Match 11.2%; Score 117; DB 20; Length 1081;

Best Local Similarity 51.8%; Pred. No. 1.5e-26;

Matches 394; Conservative 1; Mismatches 321; Indels 45; Gaps 4;

Qy 55 tgtgggcaacctgtatctaccagcgcgtgttagtggccaggaatgctgctgagggcgc 114

Db 136 tgcggcgagcgggtcatcgcgtcgcatcgtgtggtagagagacgcaactcggcgct 195

Qy 115 tggccttgccaggtcagctacactttgaccacaaactttatctatgaggttcctcgct 174

Db 196 tggccttgccagggagcctgctggtggttccaccagtatgcggagtgagcctcgct 255

Qy 175 agtgagaggttgatctgacagcagcacactgcatacaaccgacctggactacttttca 234

Db 256 agccaccctgggacactcagcggcgacactgctttgaaacctatagtagccttagtat 315

Qy 235 tatactgtgtgctaggtatctacagtagtgtagctcaaggaaacgctgtgaag----- 288

Db 316 cctccgggtgtagtgcctcaggtttggccagctgacttccatgcactccttctgagacctg 375

Qy 289 -----tactacgtgtccaaaac---gtcatccatcccaagtagtaccac 327

Db 376 caggcctactacaccgcttactctcgtatctgaaatactatctgagccctcgctacctggg 435

Qy 328 gatacaacggcagactcgcctgttgaaactgtctctcaagctcaccttcaacttctgcc 387

Db 436 aattcaccctatgacatgccttggtagcgtgtgcacctgtcactacactaacac 495

Qy 388 atcctgcctattgtgtccagtgctcaaaagcagtggtggcaattccacccctttgtgg 447

Db 496 atccagcccatctgtctccaggtccctcactattgagtttgagaacgggacagactctgg 555

Qy 448 gtgaccggtatgggaaaaagttcaagaaagttcaagtagagattaccattctgcccctcag 507

Db 556 gtgactggtggtgggtacatcaaaagaggtgagggactgccatctcccacacccctccag 615

Qy 508 gaagcagaagtagtaccattattgaccgcccaggtgttgaaacagctctcaacatccactcgg 567

Db 616 gaagttcaggtcgccatcaataaactatgtgcaacacctcttctcactgaagtagt 675

Qy 568 atcttctgagcactgagccagtcataaggaaggaagacaagaattgtgctggtgatact 627

Db 676 ttccgcaag-----gacactttggagacatggtttgtgtgccaatgcc 720

Qy 628 caaacatgaagtagtggcaagggtgattcttgaggggcctctctgctgctacattgat 687

Db 721 caaggcggaagtagctcttcgtgtagctcaggtggacccttggcctgtaacaagaat 780

Qy 688 ggtgtatgatccagacagtagtagtaagctgggattagaaatgtgtgtaaatc---tctt 744

Db 781 ggaactgtgtatcagattgtagtgcgtgagctgggagtggtggtcgcccaatcgg 840

Qy 745 cctggagctacaccaatgaatctactaccacaaatgat 785

Db 841 cccggtgtctacaccaatcagccaccactttgagtggaat 881

RESULT 9

AAx08505

ID AAA08505 standard; DNA; 1081 BP.







QY 641 atagttgcaagggtgattctggaggcctctgctgtcacattgattggtgatgcc 700  
II IIIIIII II II II II II II II II II II II II II II II II II  
Db 703 atgcttgaaggcgactcggcgccctctgctgtgctcgtgggtcagctggtgcg 762  
QY 701 agacagagtagtaagctggggattagaatgtg---gtaaatctcttctcgtggagtctaca 757  
II II II II IIIIIII II II II II II II II II II II II II II II II II II  
Db 763 agcgggggtgacagctgggtgaggtgctgcccgcaggaaccccgaggtgtctaca 822  
QY 758 ccaatgtaactactaccacaaatgattaatgccactatttcaaga 804  
II II II II IIIIIII II II II II II II II II II II II II II II II II II  
Db 823 tccgtgtcacccgccaccacaaactggatccatcggatcatcccca 869  
RESULT 13  
AAZ52473  
ID AAZ52473 standard; DNA; 1703 BP.  
XX  
AC AAZ52473;  
XX  
DT 24-FEB-2000 (first entry)  
XX  
DE HTRM clone 3376404 DNA sequence.  
XX  
KW HTRM; human transcriptional regulatory molecule; arteriosclerosis; AIDS;  
KW arteriosclerosis; cirrhosis; cancer; leukaemia; diabetes mellitus; ss;  
KW Addison's disease; multiple sclerosis; rheumatoid arthritis; infection;  
KW trauma; myasthenia gravis; adenocarcinoma; immune disorder; treatment.  
OS Homo sapiens.  
XX  
PN WO9557144-A2.  
XX  
PD 11-NOV-1999.  
XX  
PF 04-MAY-1999; 99WO-US09935.  
XX  
PR 05-MAY-1998; 98US-0084254.  
PR 07-AUG-1998; 98US-0095827.  
PR 02-OCT-1998; 98US-0102745.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Hillman JL, Bandman O, Lal P, Yue H, Reddy R, Tang YT;  
PI Gerstin EH, Patterson C, Baughn MR, Azimzai Y, Lu DAM;  
XX  
DR WPI: 2000-052941/04.  
DR P-PSDB; AAY73388.  
XX  
PT New peptides useful for diagnosis, prevention and treatment of cancer  
PT and immune disorders  
XX  
PS Claim 9; Page 190; 193pp; English.  
XX  
CC AAZ52410-252474 are human transcriptional regulator molecule (HTRM)  
CC nucleotide sequences. The HTRM protein and nucleotide sequences are  
CC useful for preventing or treating disorders associated with decreased  
CC expression or activity of HTRM which include cell proliferative  
CC disorders such as arteriosclerosis and cirrhosis; cancers including  
CC adenocarcinoma and leukaemia; immune disorders such as AIDS, Addison's  
CC disease, diabetes mellitus, rheumatoid arthritis, multiple sclerosis,  
CC systemic lupus erythematosus, and myasthenia gravis; infections and  
CC trauma. Antagonists of the HTRM polypeptides are useful for treating or  
CC preventing disorders associated with increased expression or activity of  
CC HTRMs. HTRM polypeptides, their immunogenic fragments or oligopeptides  
CC are useful for screening libraries of compounds in drug screening  
CC techniques. Polynucleotides encoding HTRM are useful for blocking the  
CC transcription of mRNA and regulating gene function by blocking the  
CC activity of HTRM. Vectors expressing HTRM or agonists can also be used to  
CC prevent or treat disorder associated with decreased HTRM expression.  
CC Antibodies which specifically bind HTRM and polynucleotides encoding HTRM  
CC are useful for diagnosing disorders associated with the expression of  
CC HTRM, particularly in assays that detect the expression of HTRM.

CC Nucleotide sequences encoding HTRM may be useful to generate  
CC hybridization probes useful in mapping the naturally occurring genomic  
CC sequence and to detect differences in gene sequences among normal,  
CC carrier and affected individuals. Using diagnostic assays, cancer can be  
CC detected prior to the appearance of clinical symptoms and thereby  
CC progression of cancer can be prevented by aggressive treatment or  
CC preventive measures.  
XX  
SQ Sequence 1703 BP; 355 A; 527 C; 528 G; 293 T; 0 other;

Query Match 11.1%; Score 115.8; DB 21; Length 1703;  
Best Local Similarity 50.6%; Pred. No. 4.5e-26;  
Matches 388; Conservative 1; Mismatches 363; Indels 15; Gaps 4;

QY 50 cagtgtgtgggcaacctgtatactccagccgctgtgtgagtgccagagatgctgctcag 109  
II IIIIIII II II II II II II II II II II II II II II II II II  
Db 705 cagcctgtgtgctcccccagatgctgaaccgaatggtggcgggcagacacgcagag 764  
QY 110 ggcgctggcctgtgcaggtgcactctttgaccacaaactttatctatggagggtccc 169  
II IIIIIII II II II II II II II II II II II II II II II II II  
Db 765 gcgagtggcctgtgcaagtgcagtcacgcgcgaacgaagccacttctcggggcgagcc 824  
QY 170 tctcagtgagaggttgatctgacagcagcagcactgcatacacccagcctgactactt 229  
II II II II IIIIIII II II II II II II II II II II II II II II II II II  
Db 825 tcctcggcgagcagtggtctcagcgctgcgactgtctccgcaacacactctgagacgt 884  
QY 230 ttcatatactgtgtgctaggtgattacagtagtgactcaaggaagaaacgtgtaagt 289  
II II II II IIIIIII II II II II II II II II II II II II II II II II II  
Db 885 cctctgaccaggtctcgtgtggggcaaggcagctgtgcagccgggaccacgcgtatgt 944  
QY 290 a---ctacgtgtccaaaatcgtcatccatcccaagtcaccaagatacaaacg-----gcag 340  
II II II II IIIIIII II II II II II II II II II II II II II II II II II  
Db 945 atcccggtgtgagcgaggtgagagcaacccctgtaccagggcagcgctccagcgctg 1004  
QY 341 acrtgccttgttgaacctgtcctcctcaagtccactctccttgcctcctcctccttattt 400  
II  
Db 1005 acgtggcctcgtgtgagctgagggcaccagtgcccttccacaaattacatctcctccgtg 1064  
QY 401 gcttgccagtgtcacaaagcagttggcaattccacccttttgggtgacggatggg 460  
II IIIIIII  
Db 1065 gcctgcctgacctcgtgtgtatttggagacgggcatgaactgctgggtcactggctggg 1124  
QY 461 gaaagttaaggaaagttcagatagagattaccattctgccttcaggaagcagaagatc 520  
II II II II IIIIIII II II II II II II II II II II II II II II II II II  
Db 1125 gcagccccagtgaggagacctcctgcgcgaacccgggactcctgcagaaactcgtgtgc 1184  
QY 521 ccattattgaccgccaggtgtgtgaacagctctacaaatcccatcgggtatcttcttgcag 580  
II  
Db 1185 ccatactgcacaccccaagtgcaacctgctctaca---gcaaaagcacacagtttggct 1241  
QY 581 cactggagccagtcacaaaggaagaaagatttgcgtgtgatactcaaaaacatgaagg 640  
II IIIIIII  
Db 1242 accaaccac 1301  
QY 641 atagttgcaagggtgattctggaggcctcctcgtgtgcattgtatggatgctatgcac 700  
II IIIIIII  
Db 1302 atgcttgaaggcgactcggcgccctcgtgtgcctcgtgggtcagtcagtcgtgcgtgc 1361  
QY 701 agacagagtagtaagctggggattagaatgtg---gtaaatctcttctcgtggagtctaca 757  
II II II II IIIIIII II II II II II II II II II II II II II II II II II  
Db 1362 agcgggggtgacagctgggtgaggtgctgcccgcaggaaccccgaggtgtctaca 1421  
QY 758 ccaatgtaactactaccacaaatgattaatgccactatttcaaga 804  
II IIIIIII  
Db 1422 tccgtgtcacccgccaccacacacacacacacacacacacacacacacacacacac 1468  
RESULT 14  
AAZ52259  
ID AAZ52259 standard; DNA; 1100 BP.  
XX  
AC AAZ52259;



Db 596 gaagtccagtcgccatcataaaactctctgtgcaaccacacctctctcctcaagtacagt 655  
QY 568 atctcttgcagcagcactggagccagtcacaaagaaagacaaagattgtgctgtgatact 627  
Db 656 ttcgcgaag-----gacattctggagacatggttctgtgcgaagcgc 700  
QY 628 caaacatgaagatagttgcaagggtgattctggaggcctctgtcgtgcacattgat 687  
Db 701 caaggcggaagatgctcttcggtgactcaggtggacccttggcctgttaacaagaat 760  
QY 688 ggtgtatgcatcacagacagagtagtaagctggggattagaatgtgtaaatc---tctt 744  
Db 761 ggaactggtatcagattggagctgagctggggagtggtgctgtgctggcgccaatcgg 820  
QY 745 cctggagctcacaccaatgaatctactaccacaaaatggat 785  
Db 821 ccgggtgtctacacaaatatacagccaccacttggatggat 861

RESULT 15  
AAF72417  
ID AAF72417 standard; cDNA; 1100 BP.  
XX AAF72417;  
AC AAF72417;  
DT 24-APR-2001 (first entry)  
XX Human PRO303 cDNA.  
DE  
XX Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;  
KW antiparkinsonian nootropic; neuroprotective; vulnery; cardiant;  
KW antiangiogenic; vasotropic; antiasthmatic; antirheumatic; cancer;  
KW antiarthritic; antinfertility; antidiabetic; antiviral; diabetes;  
KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;  
KW ischaemia; inflammation; ss.  
XX  
OS Homo sapiens.  
XX WO200104311-A1.  
PN 18-JAN-2001.  
PD 22-FEB-2000; 2000WO-US04414.  
PF 07-JUL-1999; 99US-0143048.  
PR 26-JUL-1999; 99US-0145698.  
PR 28-JUL-1999; 99US-0146222.  
PR 08-SEP-1999; 99WO-US20594.  
PR 13-SEP-1999; 99WO-US20944.  
PR 15-SEP-1999; 99WO-US21090.  
PR 15-SEP-1999; 99WO-US21547.  
PR 05-OCT-1999; 99WO-US23089.  
PR 29-NOV-1999; 99WO-US28214.  
PR 30-NOV-1999; 99WO-US28313.  
PR 16-DEC-1999; 99WO-US30095.  
PR 20-DEC-1999; 99WO-US30911.  
PR 20-DEC-1999; 99WO-US30999.  
PR 05-JAN-2000; 99WO-US00219.  
XX (GETH ) GENENTECH INC.  
PI Ashkenazi AJ, Botstein D, Desnovers L, Eaton DL, Ferrara N;  
PI Filyarov E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;  
PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kijavini LJ;  
PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;  
PI Williams PM, Wood WI;  
XX WPI: 2001-081051/09.  
DR P-PSDB; AAB80256.  
XX  
PT Sixty one nucleic acids encoding PRO polypeptides which are useful in  
the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung

PT squamous cell carcinoma) and neurodegenerative diseases (e.g.  
PT Alzheimer's disease) -  
XX  
PS Claim 2; Fig 91; 393pp; English.  
XX  
CC The present sequence is one of sixty one nucleic acids encoding novel  
CC secreted and transmembrane PRO polypeptides. The PRO polypeptides are  
CC useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung  
CC squamous cell carcinoma), gastrointestinal disorders (e.g.  
CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,  
CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.  
CC endometrial bleeding angiogenesis, ischaemia such as coronary  
CC ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,  
CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and  
CC diabetes and retinal disorders such as retinitis pigmentosum.  
CC The PRO nucleic acids have applications in molecular biology, including  
CC use as hybridization probes, and in chromosome and gene mapping.  
XX  
SQ Sequence 1100 BP; 225 A; 321 C; 314 G; 240 T; 0 other;  
  
Query Match 11.1%; Score 115.4; DB 22; Length 1100;  
Best Local Similarity 51.8%; Pred. NO. 4.8e-26;  
Matches 393; Conservative 1; Mismatches 322; Indels 45; Gaps 4;  
  
QY 55 tgtgggcaacctgtatactccagccgctgttagtggtgagcagagatgctgctgagggcgc 114  
Db 116 tgcggcgacagggtcatcacgtcgcgcatcgtggtgagagagcgcgaactggcgct 175  
QY 115 tggccttggcagcagcctacatttgaccacaaactttatctatgaggttccctcgtc 174  
Db 176 tggcgtggcagggagcctcgctggtgggttccacgtatgaggtgagcgtcgc 235  
QY 175 agtgagaggttactgacagcagacactgcatacaacccgacctggaactatttca 234  
Db 236 agcacccgtggcactcacgcgcgccactgtttgaaacctatagtgacctagtgt 295  
QY 235 tatactgtgtgctaggatcgattacagtagtgtagtgaactgaagacgtgtgaag----- 288  
Db 296 cctccgggtggatggtccagtttggccagctgacttccatgccactctctgagagcgtg 355  
QY 289 -----tactacgtgtccaaaatc---gtcatccatcccaagtaacaa 327  
Db 356 caggcctactacacccgttacttctgcatcgaatctctatctctgagccctcgtacctggg 415  
QY 328 gatacaacggcagacactgacctgttgaaactgctctcctcaagtcaactcacttctgccc 387  
Db 416 aattcacccctatgacattgacctgctgtggaagctgctgcaactgtcaactacataaac 475  
QY 388 atcctgcctatttgcctgcccagtgccacaaagcagttggcaattccaccctttgttg 447  
Db 476 atccagcccatctgtctccaggcctccacatttggatttggagacccggagacagactgt 535  
QY 448 gtgaccggatggggaaaagttaaggaaagttcagatagagattaccatttgccttcag 507  
Db 536 gtgactgctgggggtacatcaaaagagatgagggcactgcatctcccacacccctccag 595  
QY 508 gaagcagaagtaccattattgaccggcggcgttggaaacagctctataatccccatcggt 567  
Db 596 gaagttcaggtcgccatcataaacactctatgtgaaccacaccttctcctcaagtaagt 655  
QY 568 atcttctccagcactggagccagtcacaaagaaagacaaagattgtgctgtgatact 627  
Db 656 ttcgcgaag-----gacattctggagacatggttctgtgcgaagcgc 700  
QY 628 caaacatgaagatagttgcaagggtgattctggaggcctctctgtgctgcacattgat 687  
Db 701 caaggcggaagatgctcttcggtgactcaggtggacccttggcctgttaacaagaat 760  
QY 688 ggtgtatgcatcacagacagagtagtaagctggggattagaatgtgtaaatc---tctt 744  
Db 761 ggaactggtatcagattggagctgagctggggagtggtgctgtgctggcgccaatcgg 820

Qy 745 cctggagttctacaccaaataatgtaattctactaccacaaaatggat 785  
          |||||          |||||          |||||          |||||  
Db 821 cccggtgtctacaccaaataatacagccaccacttttgagtgat 861  
          |||||          |||||          |||||          |||||

Search completed: November 22, 2001, 00:22:05  
Job time: 2874 sec

This Page Blank (uspto)



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2001, 23:33:21 ; Search time 70.92 Seconds  
(without alignments)  
3324.360 Million cell updates/sec

Title: US-09-854-844-1  
Perfect score: 1041  
Sequence: 1 atgggcctgctgctgtgc.....ttattacaattgaaatga 1041

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	117	11.2	1081	4	US-09-008-271A-15
2	91.2	8.8	1097	2	US-08-978-404B-4
3	91.2	8.8	1108	2	US-09-016-366A-14
4	91.2	8.8	1108	2	US-08-978-404B-20
5	89.6	8.6	1103	2	US-09-016-366A-24
6	81.8	7.9	1031	2	US-08-978-404B-1
7	73	7.0	2038	4	US-09-008-271A-18
8	72.8	7.0	1137	2	US-09-016-366A-18
9	72.8	7.0	1137	2	US-08-978-404B-13
10	72.8	7.0	1154	2	US-09-016-366A-16
11	72.8	7.0	1154	2	US-08-978-404B-11
12	72.2	6.9	1219	2	US-08-978-404B-7
13	71.2	6.8	735	4	US-09-079-970A-1
14	71.2	6.8	771	4	US-09-079-970A-4
15	71.2	6.8	1081	2	US-09-016-366A-22
16	71.2	6.8	1081	2	US-08-978-404B-17
17	71.2	6.8	1128	2	US-09-016-366A-20
18	71.2	6.8	1128	2	US-08-978-404B-15
19	68.2	6.6	1109	4	US-09-088-651-6
20	66.4	6.4	1109	4	US-09-088-651-1
21	62	6.0	821	1	US-08-650-129-2
22	62	6.0	821	3	US-08-984-417-2
23	62	6.0	866	3	US-08-650-129-3
24	62	6.0	866	3	US-08-984-417-3
25	57.2	5.5	2413	4	US-09-518-046-1
26	56.8	5.5	696	1	US-08-508-448C-24
27	56.8	5.5	1517	1	US-08-508-448C-15

28	56	5.4	1605	2	US-09-000-846-1	Sequence 1, Appl
29	55.8	5.4	1065	1	US-08-427-640-1	Sequence 1, Appl
30	55.8	5.4	1065	1	US-08-427-640-5	Sequence 5, Appl
31	55.8	5.4	1068	1	US-08-427-640-3	Sequence 3, Appl
32	55.8	5.4	1955	2	US-08-883-795A-39	Sequence 39, Appl
33	55.8	5.4	2457	6	5344773-1	Patent No. 5344773
34	55.8	5.4	7360	1	US-08-286-740-1	Sequence 1, Appl
35	55.8	5.4	7360	5	PCT-US95-09576-1	Sequence 1, Appl
36	55.6	5.3	2416	4	US-09-261-416-1	Sequence 1, Appl
37	54.8	5.3	1077	3	US-08-807-151-2	Sequence 2, Appl
38	54.8	5.3	1479	4	US-09-342-749-1	Sequence 1, Appl
39	54.4	5.2	1095	2	US-08-978-404B-9	Sequence 9, Appl
40	54.4	5.2	1186	4	US-09-008-271A-17	Sequence 17, Appl
41	54.2	5.2	453	6	5200340-3	Patent No. 5200340
42	54.2	5.2	1065	2	US-08-811-949-60	Sequence 60, Appl
43	54.2	5.2	1068	1	US-08-137-116-2	Sequence 2, Appl
44	54.2	5.2	1068	1	US-08-427-640-7	Sequence 7, Appl
45	54.2	5.2	1068	2	US-08-811-949-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1  
US-09-008-271A-15  
; Sequence 15, Application US/09008271A  
; Patent No. 6203979  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; Hillman, Jennifer L.  
; Yue, Henry  
; Guegler, Karl J.  
; Corley, Neil C.  
; Tang, Tom Y.  
; Shah, Purvi  
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/008,271A  
; FILING DATE: 16-Jan-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: <Unknown>  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mohan-Peterson, Sheela  
; REGISTRATION NUMBER: 41,201  
; REFERENCE/DOCKET NUMBER: PF-0458 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1081 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: PROSTUT03  
; CLONE: 789927  
; SEQUENCE DESCRIPTION: SEQ ID NO: 15 :  
US-09-008-271A-15

by  
protease activity  
homology

```
Query Match      11.2%; Score 117; DB 4; Length 1081;
Best Local Similarity 51.8%; Pred. No. 2.8e-29;
Matches 394; Conservative 1; Mismatches 321; Indels 45; Gaps 4;

QY 55 tgggcaactctatactccagccgctgttaggtggcagagatgctctgagggcgc 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 TGGCGCGGACGGGTATCATCAGTCGCGCATCGTGGGTGGAGAGAGCCCGCACTCGGGCGT 195

QY 115 tggcctggcaggtcagctacactttgaccacaactttatctatgaggggtccctcgtc 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 TGGCCGTGGCAGGGAGCCTGCGCCTGTGGGATTCCACGATATGCGAGTGTAGCCTGCTC 255

QY 175 agtgagaggttgatactgacagacagacactgcatacaaccgacctggactacttttca 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 AGCCACCGCTGGGCACATCAGCGGGCGCACGTGCTTTGAAACCTATTAGTGACCTTAGTGAT 315

QY 235 tatactgtgtggctagatcgattacagtagtgactcaaggaaacggtgtgaag----- 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 CCCTCGGGTGGATGTCAGTTTGGCCAGCTGACTTCCATGCCATCCTTCTCGAGCCTG 375

QY 289 -----tactcgtgtccaaaatc---gtccatccatcccaagttaccaa 327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 CAGGCTACTACACCGCTTACTTCGTATCGAATATCTATCTGAGCCCTCGCTACCTGGGG 435

QY 328 gatacaacggcagacactcgtcctgttgaaactcctcctcaagtcaacctcactctgcc 387
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 AATCACCCCTATGACATTCGCTTTGGTGAGCTGTGACCTGTGACCTACACTAANAC 495

QY 388 atctcgtctatttgcctggccaggtgcaaaagcagttggcattccaccctttgttgg 447
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 ATCCAGGCCATCTGTCTCCAGGCTCCACATTTGAGTTTGAGAACGGACAGACTGTGG 555

QY 448 gtacocgga tggggaaagttagggaaagtctcagatagagattaccattctgccttcag 507
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 556 GTGACTGGCTGGGGGTACATCAAGAGGATGAGGCATGCGCATCTCCCCACACCTCCAG 615

QY 508 gaagcagaataccattatgtaccgcaggtctgtgaacagctctacaatcccatcgt 567
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 616 GAAGTTGAGTGGCGCATTAACAACTCTATGTGCNAACCACTCTTCCCTCAAGTACAGT 675

QY 568 atcttctgcagcactgggagcagtcactcaaggaaagacagatttggctggtgatact 627
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 676 TTCGCGAAG-----GACATCTTTGGAGACATGGTTTGTGCTGGCAATGCC 720

QY 628 caaaacataagagatagttgcaaggggtgattcttgaggggacctctgctgctcattgat 687
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 CAAGCGGGAAGGATGCTGCTGCGGTGACTCAGGTGGACCTTGGCCTGTGTAACAAGAA 780

QY 688 ggtgtatggatccagacagagtagtagtgaagctggggttagaagttagaataatc---tctt 744
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 781 GGACTGTGGTATCATCAGATGGAGTGGTGGAGTGGGGAGTGGGCTGTGGTGGCCCAATCGG 840

QY 745 cctggagctacacccaagttaattactactaccacaaaatggat 785
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 841 CCCGGTGTCTACACCAATATCAGGCCACCACTTTCAGTGGAT 881
```

```
RESULT 2
US-08-978-404B-4
; Sequence 4, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
```

```
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1097 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-978-404B-4
```

```
Query Match      8.8%; Score 91.2; DB 2; Length 1097;
Best Local Similarity 50.4%; Pred. No. 1.3e-20;
Matches 365; Conservative 1; Mismatches 334; Indels 24; Gaps 5;

QY 80 gcgttgtagtgccagagatgctgctcagggcgctcaggcctgttgccaggtcagctcacact 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 GCATTGTGGGAGGACGAGAGGCTTCTGAAAGTAAGTGGCCCTGCGAGGTGAGCTGAGAT 170

QY 140 tggaccacaact-----tatctatggaggttccctcgtcagtgagaggttgatcac 190
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 TTAATTCAGCTTCTGGATGCATTTCTGTGGCGCTCCCTCATTCACCCACAGTGGGTGC 230

QY 191 tgacagcagcacactgcatacaaccgacctggacta---cttttctatatactgtgtggc 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 231 TCACTGCGGCACACTGTGTGGACTGACATCAAAAGCCACAGACTCTTCCGTGTACAGC 290

QY 248 taggatgattacagtaggtgactcaaggaaacgtggaagtactacgtgccaaaatcgc 307
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291 TTCGTGAGCAGTATCTATACTATGCGGACCACTGACTGTGAACCGGACCGTGTGTGC 350

QY 308 tcattccatcccaagtaccagaatacaacggcagacrtcgcttgttgaacctgtccttc 367
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 351 ACCCCCACTACTACAGTCGAGGATGGGCGAGACATTTGCCCTGCTGGAGCTTGAGATCC 410

QY 368 aagtcaccttcaactctgcatcctgcctctatttgcctgcccagtgctcaaaagcagttgg 427
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 CTGTGAATGTCTCCACCCATATCCACCCCATATCCTGCCCCCTGCTCGGAGACCTTCC 470

QY 428 caattccaccctttgttgggtgaccggatgggaaagttaaagaaagttcagatagag 487
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 CCTCGGGACTTCTTGTGGGTAACAGGCTGGGGCGACATTGATAGTACAGAGCCTCTCC 530

QY 488 attaccattctgccttcaggaaacagaagtaccattattgaccgcgaggtcttgtgaac 547
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 TGCACACCTTATCTCTGAAGCAAGTGAAGGTCCCCATTTGTGAAAACACAGCTGTGTGATC 590

QY 548 agctctacaatcccatcggtatctcttcttgcagcactggagccagtcatacgaagaagaca 607
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 GGAAGTAC---CACACTGGCCTCTACACAGGAGATGATGTTCCCATTTGCCAGGATGCA 647

QY 608 agatttgtgtggtgatactcaaacatgaaggatagttgcaagggtgattcttgagggc 667
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```









Db 765 TACACCGTGTCACTACTACTTGGACTGGAT 796

RESULT 9  
US-08-978-404B-13  
; Sequence 13, Application US/08978404B  
; Patent No. 5968782  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
; TITLE OF INVENTION: FIBRINOGEN  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,404B  
; FILING DATE: 25-NOV-97  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/032,354  
; FILING DATE: 04-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7090  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1137 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-978-404B-13

Query Match 7.0%; Score 72.8; DB 2; Length 1137;  
Best Local Similarity 52.2%; Pred. No. 2.1e-14;  
Matches 236; Conservative 1; Mismatches 203; Indels 12; Gaps 3;  
QY 337 gcagactgcctgttgaaactgtcctcgaagtcaccttcactctgtccactcgtcct 396  
Db 354 GCGGACATGCCCTGCTGGAGCTGGAGGCGGTGAACGTCTCCAGCCACGCTCCACG 413  
QY 397 atttgcctgcctgtgtcacaagcagttggcattccaccctttgttggtgacgga 456  
Db 414 GTACCCCTGCCCTGCTCAGAGACCTTCCCCCGGGGATGCCGTCTGGGTCACTGGC 473  
QY 457 tggggaaggttaagaaagttcagatagagattaccattctgccttcagggaagcagaa 516  
Db 474 TGGGGCGATGTGGACATGATGAGCGGCTCCACCGCCATTTCTCTGAGAGAGGTGAAG 533  
QY 517 gtaccattattgacgcgcaggttbtgaaagctctacaatcccatcggtatcttctg 576  
Db 534 GTCCCCATAATGGAACCAACATTTGTGACGCAAAATACCACC---TTGGCGCTACACG 590  
QY 577 ccagcactggagccagtcacaggaagacaagatttgcgtgtgatactcaaacatg 636  
Db 591 GGAGACAGCTGCCGATCGTCCGTGAGACATGCTGTGTGCCGGGAACACCCG-----G 644

QY 637 aaggatagttgcaagggtgattcttgaggggcctctgtcgtgtcacattgatgtgtatgg 696  
Db 645 AGGACTCATGCGCAGGCGACTCCGGAGGGCCCTGCTGTGCAAGGTGAATGGCACCTGG 704  
QY 697 atccagacaggtagtagtaagctggggattagaatgtggtaaatc---tcttctcgagtc 753  
Db 705 CTCAGCGGGCGGTGGTCACTGGGCGAGGGCTGTGCCAGCCCAACCGGCTGGCATC 764  
QY 754 tacacaatgtatctactaccacaaatggat 785  
Db 765 TACACCGTGTCACTACTACTTGGACTGGAT 796

RESULT 10  
US-09-016-366A-16  
; Sequence 16, Application US/09016366A  
; Patent No. 5955431  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; APPLICANT: Huang, Chifu  
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
; TITLE OF INVENTION: INHIBITORS  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,366A  
; FILING DATE: January 30, 1998  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/037,090  
; FILING DATE: 05-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7093  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1154 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-09-016-366A-16

Query Match 7.0%; Score 72.8; DB 2; Length 1154;  
Best Local Similarity 52.2%; Pred. No. 2.1e-14;  
Matches 236; Conservative 1; Mismatches 203; Indels 12; Gaps 3;  
QY 337 gcagactgcctgttgaaactgtcctcgaagtcaccttcactctgtccactcgtcct 396  
Db 375 GCGGATATGCCCTGCTGGAGCTGGAGGCGGTGAACATCTCCAGCCGCTCCACAG 434  
QY 397 atttgcctgcctgtgtcacaagcagttggcattccaccctttgttggtgacgga 456  
Db 435 GTACATGCTGCCCTGCTCGGAGACCTTCCCCCGGGGATGCCGTCTGGGTCACTGGC 494  
QY 457 tggggaaggttaagaaagttcagatagagattaccattctgccttcagggaagcagaa 516

Db 495 TGGGGCGATGTGGACAATGATGAGCCCTCCACCGCCATTTCCCTGGAAGCAGGTGAAG 554  
QY 517 gtaccattattgacgcagcaggttgaacagctctacaatcccatcggtatctcttg 576  
Db 555 GTCCCATATATGAAAACACATTTGTGACGCAAAATACCAAC---TTGGCGCCTACACG 611  
QY 577 ccagcactgagccagtcacaaaggaagcaagattgtgctggtgatactcaaaacatg 636  
Db 612 GGAGACGCTCCGCATCATCCGTGACGACATGCTGTGTCGCGGGAACAGCCA-----G 665  
QY 637 aagtagttgcaaggtgattctgagggccctctctgctgtcacattgatggtgatg 696  
Db 666 AGGACTCCTGCAAGGCGGACTCTGGAGGCGCCCTGTGTGTGCAAGGTGAATGGCACCTGG 725  
QY 697 atccagacagagtagtaagctaggggattagaatggtgataatc---tcttctgagtc 753  
Db 726 CTACAGCGGCGGTGTGTGACGTGGGAGGCGGTGTGTGTCGCGGGAACAGCCA-----G 785  
QY 754 tacaccaatgtaactactactacacaaatggat 785  
Db 786 TACACCGGTGCTACCTACTACTTGGACTGGAT 817

RESULT 11  
US-08-978-404B-11  
; Sequence 11, Application US/08978404B  
; Patent No. 5968782  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
; TITLE OF INVENTION: FIBRINOGEN  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA: US/08/978,404B  
; FILING DATE: 25-NOV-97  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER: 60/032,354  
; FILING DATE: 04-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7090  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1154 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-978-404B-11

Query Match 7.08; Score 72.8; DB 2; Length 1154;  
Best Local Similarity 52.28; Pred. No. 2.1e-14;  
Matches 236; Conservative 1; Mismatches 203; Indels 12; Gaps 3;

QY 337 gagacrtcgccctgttgaaactgtctctcaagtcaactcactcttctgcccctgcgcct 396  
Db 375 GCGGATATGCCCTGCTGGAGCTGGAGGAGCCGCGTGAACATCTCCAGCCGCGTCCACAG 434  
QY 397 atttcttccagtgatcacaagcaggttggcaattccacccttttgggtgacccgga 456  
Db 435 GTATGCTGCCCCCTGCTCGGAGACCTTCCCCCGGGGATGCCGTGCTGGGTCACTGGC 494  
QY 457 tggggaaaaagtttaagaaagttcagatagattacacattctgccccttcagggaagcagaa 516  
Db 495 TGGGGCGATGTGGACAATGATGAGCCCTCCACCGCCATTTCCCTGGAAGCAGGTGAAG 554  
QY 517 gtaccattattgacgcagcaggttgaacagctctacaatcccatcggtatctcttg 576  
Db 555 GTCCCATATATGAAAACACATTTGTGACGCAAAATACCAAC---TTGGCGCCTACACG 611  
QY 577 ccagcactgagccagtcacaaaggaagcaagattgtgctggtgatactcaaaacatg 636  
Db 612 GGAGACGCTCCGCATCATCCGTGACGACATGCTGTGTCGCGGGAACAGCCA-----G 665  
QY 637 aagtagttgcaaggtgattctgagggccctctctgctgtcacattgatggtgatg 696  
Db 666 AGGACTCCTGCAAGGCGGACTCTGGAGGCGCCCTGTGTGTGCAAGGTGAATGGCACCTGG 725  
QY 697 atccagacagagtagtaagctaggggattagaatggtgataatc---tcttctgagtc 753  
Db 726 CTACAGCGGCGGTGTGTGACGTGGGAGGCGGTGTGTGTCGCGGGAACAGCCA-----G 785  
QY 754 tacaccaatgtaactactactacacaaatggat 785  
Db 786 TACACCGGTGCTACCTACTACTTGGACTGGAT 817

RESULT 12  
US-08-978-404B-7  
; Sequence 7, Application US/08978404B  
; Patent No. 5968782  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
; TITLE OF INVENTION: FIBRINOGEN  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA: US/08/978,404B  
; FILING DATE: 25-NOV-97  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER: 60/032,354  
; FILING DATE: 04-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7090  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1219 base pairs  
; TYPE: nucleic acid





## Gaps

```

QY 337 gcagacrtcgccctgttgaaactgtcctcctcaagtcacaccttcacattcttgccatcctgcct 396
Db 335 GCGACATCGCCCTGCTGGAGCTGGAGGAGCGGTGAAGGTCTCCAGCCACGTCACACG 394
QY 397 attgttgccagtgatcaacaagcggttggaattccacccttttgggtgacccgga 456
Db 395 GTCACCTGCCCCCTGCCCTCAGAGACCTTCCCCCGGGGATGCCGTGCTGGGTCACTGGC 454
QY 457 tggggaaggttaaggaagttcagatagagattaccattctgcccttcaggaagcagaa 516
Db 455 TGGGGCATGTGACAAATGATGAGCGCCTCCACCGCCATTTCCTCTGAAGCAGGTGAAG 514
QY 517 gtaccattattgacccgaggttggaacagctctacaaatcccatcggtatctctttg 576
Db 515 GTCCCATATAATGGAACACCAATTTGTGACGCAAAATACCACG---TTGGCGCTACACG 571
QY 577 ccagcactgagccagtcacaaaggaagaaagatttggctggtgatactcaaaacatg 636
Db 572 GGAGACGACGTCCGCATCGTCCGTGACGACATGCTGTGTCGGGGAACACCCG-----G 625
QY 637 aaggatagttgcaagggtgattctgagggcctctctgctgtcacattgatgtgatgg 696
Db 626 AGGACTCATGCCCAGGGCGACTCCGGAGGGCCCCCTGGTGTGCAAGGTGAATGGACCTGG 685
QY 697 atccagacaggagtagtaagctgggattagaaatgtggtaaatc---tcttcctggagtc 753
Db 686 CTGAGCGGGCGTGGTCACTGCGGCGAGGGCTGTGCCCCAGCCCAACCGCCCTGGCATC 745
QY 754 tacaccaatgtaatctactactaccataaatggat 785
Db 746 TACACCGGTCTACCTACTACTTGGACTGGAT 777

```

Search completed: November 22, 2001, 00:19:42  
Job time: 2781 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2001, 00:21:21 ; Search time 1439.33 Seconds  
(without alignments)  
11187.099 Million cell updates/sec

Title: US-09-854-844-1  
Perfect score: 1041  
Sequence: 1 atgggccctgctgctgc.....ttattacaatttgaaatga 1041

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_ba3.\*
- 4: gb\_in1.\*
- 5: gb\_in2.\*
- 6: gb\_in3.\*
- 7: gb\_om.\*
- 8: gb\_ov.\*
- 9: gb\_pat1.\*
- 10: gb\_pat2.\*
- 11: gb\_ph.\*
- 12: gb\_pl1.\*
- 13: gb\_pl2.\*
- 14: gb\_pl3.\*
- 15: gb\_pl4.\*
- 16: en\_ba1.\*
- 17: en\_ba2.\*
- 18: en\_fun.\*
- 19: en\_htgo\_hum.\*
- 20: en\_htgo\_inv.\*
- 21: en\_htgo\_rod.\*
- 22: en\_htg\_hum1.\*
- 23: en\_htg\_hum2.\*
- 24: en\_htg\_hum3.\*
- 25: en\_htg\_hum4.\*
- 26: en\_htg\_hum5.\*
- 27: en\_htg\_hum6.\*
- 28: en\_htg\_hum7.\*
- 29: en\_htg\_hum8.\*
- 30: en\_htg\_inv1.\*
- 31: en\_htg\_inv2.\*
- 32: en\_htg\_other.\*
- 33: en\_htg\_rod.\*
- 34: en\_hum1.\*
- 35: en\_hum2.\*
- 36: en\_hum3.\*
- 37: en\_hum4.\*
- 38: en\_hum5.\*
- 39: en\_hum6.\*
- 40: en\_hum7.\*
- 41: en\_in.\*
- 42: en\_om.\*
- 43: en\_or.\*

- 44: em\_ov.\*
- 45: em\_pat.\*
- 46: em\_ph.\*
- 47: em\_pl.\*
- 48: em\_ro.\*
- 49: em\_sts.\*
- 50: em\_sy.\*
- 51: em\_un.\*
- 52: em\_vi.\*
- 53: gb\_sts1.\*
- 54: gb\_sts2.\*
- 55: gb\_sts3.\*
- 56: gb\_sy.\*
- 57: gb\_un.\*
- 58: gb\_vil.\*
- 59: gb\_vil2.\*
- 60: gb\_htg1.\*
- 61: gb\_htg2.\*
- 62: gb\_htg3.\*
- 63: gb\_htg4.\*
- 64: gb\_htg5.\*
- 65: gb\_htg6.\*
- 66: gb\_htg7.\*
- 67: gb\_htg8.\*
- 68: gb\_htg9.\*
- 69: gb\_htg10.\*
- 70: gb\_htg11.\*
- 71: gb\_htg12.\*
- 72: gb\_htg13.\*
- 73: gb\_htg14.\*
- 74: gb\_htg15.\*
- 75: gb\_htg16.\*
- 76: gb\_htg17.\*
- 77: gb\_htg18.\*
- 78: gb\_htg19.\*
- 79: gb\_htg20.\*
- 80: gb\_htg21.\*
- 81: gb\_htg22.\*
- 82: gb\_htg23.\*
- 83: gb\_htg24.\*
- 84: gb\_htg25.\*
- 85: gb\_pr1.\*
- 86: gb\_pr2.\*
- 87: gb\_pr3.\*
- 88: gb\_pr4.\*
- 89: gb\_pr5.\*
- 90: gb\_pr6.\*
- 91: gb\_pr7.\*
- 92: gb\_pr8.\*
- 93: gb\_pr9.\*
- 94: gb\_rol.\*
- 95: gb\_rol2.\*
- 96: gb\_in4.\*
- 97: gb\_pr10.\*
- 98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB ID	Description
		Match	Length	%		
C 1	290	27.9	178181	77	AC084420	AC084420 Homo sapi
2	290	27.9	180155	70	AC026556	AC026556 Homo sapi
3	215	20.7	178181	77	AC084420	AC084420 Homo sapi
C 4	52	5.0	180155	70	AC026556	AC026556 Homo sapi
5	21	2.0	122280	85	AC004847	AC004847 Homo sapi
C 6	21	2.0	159468	63	AC015280	AC015280 Drosophila
C 7	21	2.0	170282	4	AC006467	AC006467 Drosophila
8	21	2.0	252527	77	AC084407	AC084407 Mus muscu

```
c 9 21 2.0 311800 5 AE003781
10 20 1.9 2225 1 AF076604
11 20 1.9 12018 1 AF121254
12 20 1.9 16098 94 MM063418
13 20 1.9 66494 86 AC006957
14 20 1.9 59855 87 AC018760
15 20 1.9 112622 61 AC010866
16 20 1.9 113254 89 AL137009
17 20 1.9 122961 93 HSJ193N13
18 20 1.9 123397 85 AC004808
19 20 1.9 134222 85 AB045363
20 20 1.9 146358 92 HS3D11
21 20 1.9 146463 73 AC068306
22 20 1.9 150152 79 AL355502
23 20 1.9 152719 62 AC012545
24 20 1.9 153487 81 AL513208
25 20 1.9 156880 86 AC006466
26 20 1.9 159235 72 AC055782
27 20 1.9 169542 80 AL356482
28 20 1.9 172148 74 AC073132
29 20 1.9 180742 90 AL359753
30 20 1.9 181505 67 AC022254
31 20 1.9 184306 73 AC068547
32 20 1.9 188603 76 AC079790
33 20 1.9 195418 68 AC023524
34 20 1.9 200000 91 AP000493
35 20 1.9 292721 89 AF130343
36 20 1.9 300000 91 AP002534
37 20 1.9 347550 2 AP001118
38 20 1.8 837 53 CNS06WMC
39 19 1.8 888 53 CNS06UMH
40 19 1.8 891 53 CNS06UPZ
41 19 1.8 904 53 CNS06US6
42 19 1.8 1003 53 CNS06XUJ
43 19 1.8 1064 53 CNS06XUI
44 19 1.8 2655 5 AF294795
45 19 1.8 3564 94 AF035151
```

ALIGNMENTS

```
RESULT 1
AC084420/c
LOCUS
DEFINITION Homo sapiens chromosome RPC1-11 clone RP11-415A13, WORKING DRAFT
SEQUENCE, 32 unordered pieces.
AC084420
VERSION AC084420.3 GI:11612633
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 178181)
Waterston,R.H.
The sequence of Homo sapiens clone
2 (bases 1 to 178181)
Waterston,R.H.
Direct Submission
Submitted (03-NOV-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Dec 11, 2000 this sequence version replaced gi:11136874.
```

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0415A13
----- Summary Statistics -----
```

```
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 160877 bases at least Q40
Consensus quality: 166793 bases at least Q30
Consensus quality: 169458 bases at least Q20
Insert size: 178000; agarose-fp
Quality coverage: 3.40 in Q20 bases; agarose-fp
Quality coverage: 3.71 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
```

```
1 1401: contig of 1401 bp in length
1402 1501: gap of unknown length
1502 2671: contig of 1170 bp in length
2672 2771: gap of unknown length
2772 4499: contig of 1728 bp in length
4500 4599: gap of unknown length
4600 6309: contig of 1710 bp in length
6310 6409: gap of unknown length
6410 8601: contig of 2192 bp in length
8602 8701: gap of unknown length
8702 10415: contig of 1714 bp in length
10416 10515: gap of unknown length
10516 13064: contig of 2549 bp in length
13065 13164: gap of unknown length
13165 16156: contig of 2992 bp in length
16157 16256: gap of unknown length
16257 18638: contig of 2382 bp in length
18639 18739: gap of unknown length
18739 21463: contig of 2724 bp in length
21463 21562: gap of unknown length
21563 24282: contig of 2720 bp in length
24283 24382: gap of unknown length
24383 28368: contig of 3986 bp in length
28369 28468: gap of unknown length
28469 32249: contig of 3781 bp in length
32250 32349: gap of unknown length
32350 36555: contig of 4206 bp in length
36556 36655: gap of unknown length
36656 42852: contig of 6197 bp in length
42853 42952: gap of unknown length
42953 47243: contig of 4291 bp in length
47244 47343: gap of unknown length
47344 53015: contig of 5672 bp in length
53016 53115: gap of unknown length
53116 59822: contig of 6707 bp in length
59823 59922: gap of unknown length
59923 65341: contig of 5419 bp in length
65342 65441: gap of unknown length
65442 70898: contig of 5457 bp in length
70899 70998: gap of unknown length
70999 76921: contig of 5923 bp in length
76922 77021: gap of unknown length
77022 85604: contig of 8583 bp in length
85605 85704: gap of unknown length
85705 90872: contig of 5168 bp in length
90873 90973: contig of 6343 bp in length
90973 97316: gap of unknown length
97316 105928: contig of 8513 bp in length
105929 106028: gap of unknown length
106029 116153: contig of 10125 bp in length
116154 116253: gap of unknown length
```



Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission  
 Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On May 30, 2000 this sequence version replaced gi:7652030.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L7426

Center clone name: 731\_D\_1

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 158918 bases at least Q40

Consensus quality: 169606 bases at least Q30

Consensus quality: 174264 bases at least Q20

Insert size: 182000; agarose-fp

Insert size: 177255; sum-of-contigs

Quality coverage: 3.7 in Q20 bases; agarose-fp

Quality coverage: 3.8 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 30 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 1118: contig of 1118 bp in length  
 \* 1119 1218: gap of 100 bp  
 \* 1219 2286: contig of 1068 bp in length  
 \* 2287 2386: gap of 100 bp  
 \* 2387 3683: contig of 1297 bp in length  
 \* 3684 3783: gap of 100 bp  
 \* 3784 5036: contig of 1253 bp in length  
 \* 5037 5136: gap of 100 bp  
 \* 5137 6185: contig of 1049 bp in length  
 \* 6186 6285: gap of 100 bp  
 \* 6286 7576: contig of 1291 bp in length  
 \* 7577 7676: gap of 100 bp  
 \* 7677 8942: contig of 1266 bp in length  
 \* 8943 9042: gap of 100 bp  
 \* 9043 9621: contig of 579 bp in length  
 \* 9622 9721: gap of 100 bp  
 \* 9722 11160: contig of 1439 bp in length  
 \* 11161 11260: gap of 100 bp  
 \* 11261 13550: contig of 2290 bp in length  
 \* 13551 13650: gap of 100 bp  
 \* 13651 17467: contig of 3817 bp in length  
 \* 17468 17567: gap of 100 bp  
 \* 17568 20973: contig of 3406 bp in length  
 \* 20974 21073: gap of 100 bp

\* 21074 23243: contig of 2170 bp in length  
 \* 23244 23343: gap of 100 bp  
 \* 23344 26526: contig of 3183 bp in length  
 \* 26527 26626: gap of 100 bp  
 \* 26627 31028: contig of 4402 bp in length  
 \* 31029 31128: gap of 100 bp  
 \* 31129 38352: contig of 7224 bp in length  
 \* 38353 38452: gap of 100 bp  
 \* 38453 43015: contig of 4563 bp in length  
 \* 43016 43115: gap of 100 bp  
 \* 43116 49226: contig of 6111 bp in length  
 \* 49227 49326: gap of 100 bp  
 \* 49327 54582: contig of 5256 bp in length  
 \* 54583 54682: gap of 100 bp  
 \* 54683 60160: contig of 5478 bp in length  
 \* 60161 60260: gap of 100 bp  
 \* 60261 66143: contig of 5883 bp in length  
 \* 66144 66243: gap of 100 bp  
 \* 66244 74231: contig of 7908 bp in length  
 \* 74232 74331: gap of 100 bp  
 \* 74332 80835: contig of 6504 bp in length  
 \* 80836 80935: gap of 100 bp  
 \* 80936 91055: contig of 10120 bp in length  
 \* 91056 91155: gap of 100 bp  
 \* 91156 101763: contig of 10608 bp in length  
 \* 101764 101863: gap of 100 bp  
 \* 101864 113795: contig of 11932 bp in length  
 \* 113796 113895: gap of 100 bp  
 \* 113896 125368: contig of 11473 bp in length  
 \* 125369 125468: gap of 100 bp  
 \* 125469 145206: contig of 19738 bp in length  
 \* 145207 145306: gap of 100 bp  
 \* 145307 162142: contig of 16836 bp in length  
 \* 162143 162242: gap of 100 bp  
 \* 162243 180155: contig of 17913 bp in length.

FEATURES Location/Qualifiers

source  
 1..180155  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4"  
 /clone="RP11-731D1"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 misc\_feature  
 1..1118  
 /note="assembly\_fragment"  
 misc\_feature  
 1219..2286  
 /note="assembly\_fragment"  
 misc\_feature  
 2387..3683  
 /note="assembly\_fragment"  
 misc\_feature  
 3784..5036  
 /note="assembly\_fragment"  
 misc\_feature  
 5137..6185  
 /note="assembly\_fragment"  
 misc\_feature  
 6286..7576  
 /note="assembly\_fragment"  
 misc\_feature  
 7677..8942  
 /note="assembly\_fragment"  
 misc\_feature  
 9043..9621  
 /note="assembly\_fragment"  
 misc\_feature  
 clone\_end:T7  
 vector\_side:right  
 9722..11160  
 /note="assembly\_fragment"  
 misc\_feature  
 11261..13550  
 /note="assembly\_fragment"  
 misc\_feature  
 13651..17467  
 /note="assembly\_fragment"  
 misc\_feature  
 17568..20973  
 /note="assembly\_fragment"  
 misc\_feature  
 21074..23243  
 /note="assembly\_fragment"  
 misc\_feature  
 23344..26526  
 /note="assembly\_fragment"



KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	Waterston, R.H.
JOURNAL	The sequence of Homo sapiens clone
REFERENCE	1 (bases 1 to 178181)
AUTHORS	Waterston, R.H.
TITLE	2 (bases 1 to 178181)
JOURNAL	Direct Submission
COMMENT	Submitted (03-NOV-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
	On Dec 11, 2000 this sequence version replaced gi:11136874.
	----- Genome Center -----
	Center: Washington University Genome Sequencing Center
	Center code: WUGSC
	Web site: http://genome.wustl.edu/gsc/index.shtml
	----- Project Information -----
	Center project name: H_NH0415A13
	----- Summary Statistics -----
	Sequencing vector: M13; 100%
	Chemistry: Dye-primer ET; 100% of reads
	Chemistry: Dye-terminator Big Dye; 0% of reads
	Assembly program: Phrap; version 0.990319
	Consensus quality: 160877 bases at least Q40
	Consensus quality: 166793 bases at least Q30
	Consensus quality: 169458 bases at least Q20
	Insert size: 188000; agarose-fp
	Insert size: 175081; sum-of-contigs
	Quality coverage: 3.40 in Q20 bases; agarose-fp
	Quality coverage: 3.71 in Q20 bases; sum-of-contigs
	-----
	* NOTE: This is a 'working draft' sequence. It currently
	* consists of 32 contigs. The true order of the pieces
	* is not known and their order in this sequence record is
	* arbitrary. Gaps between the contigs are represented as
	* runs of N, but the exact sizes of the gaps are unknown.
	* This record will be updated with the finished sequence
	* as soon as it is available and the accession number will
	* be preserved.
	-----
	* 1401: contig of 1401 bp in length
	* 1402: gap of unknown length
	* 1502: contig of 1170 bp in length
	* 2672: gap of unknown length
	* 2772: contig of 1728 bp in length
	* 4500: gap of unknown length
	* 4600: contig of 1710 bp in length
	* 6310: gap of unknown length
	* 6410: contig of 2192 bp in length
	* 8602: gap of unknown length
	* 8702: contig of 1714 bp in length
	* 10416: gap of unknown length
	* 10516: contig of 2549 bp in length
	* 13065: gap of unknown length
	* 13165: contig of 2992 bp in length
	* 16157: gap of unknown length
	* 16256: contig of 2382 bp in length
	* 16529: gap of unknown length
	* 18638: contig of 2382 bp in length
	* 18739: gap of unknown length
	* 21462: contig of 2724 bp in length
	* 21562: gap of unknown length
	* 21563: contig of 2720 bp in length
	* 24283: gap of unknown length
	* 24383: contig of 3986 bp in length
	* 28369: gap of unknown length
	* 28468: contig of 3781 bp in length
	* 28469: gap of unknown length
	* 32249: contig of 4206 bp in length
	* 32350: gap of unknown length
	* 36555: contig of 4206 bp in length
	* 36556: gap of unknown length

\* 36656 42852: contig of 6197 bp in length  
\* 42853 42952: gap of unknown length  
\* 42953 47243: contig of 4291 bp in length  
\* 47244 47343: gap of unknown length  
\* 47344 53015: contig of 5672 bp in length  
\* 53016 53115: gap of unknown length  
\* 53116 59022: contig of 6707 bp in length  
\* 59023 59223: gap of unknown length  
\* 59224 65341: contig of 5419 bp in length  
\* 65342 65442: gap of unknown length  
\* 65443 70998: contig of 5457 bp in length  
\* 70999 70998: gap of unknown length  
\* 70999 76921: contig of 5923 bp in length  
\* 76922 77021: gap of unknown length  
\* 77022 85604: contig of 8583 bp in length  
\* 85605 85704: gap of unknown length  
\* 85705 90872: contig of 5188 bp in length  
\* 90873 90873: gap of unknown length  
\* 90873 97315: contig of 6343 bp in length  
\* 97316 97415: gap of unknown length  
\* 97416 105928: contig of 8513 bp in length  
\* 105929 106028: gap of unknown length  
\* 106030 116153: contig of 10125 bp in length  
\* 116154 116253: gap of unknown length  
\* 116254 127251: contig of 10998 bp in length  
\* 127252 127351: gap of unknown length  
\* 127352 136839: contig of 9488 bp in length  
\* 136840 136939: gap of unknown length  
\* 136940 142282: contig of 5343 bp in length  
\* 142283 142382: gap of unknown length  
\* 142383 149737: contig of 7355 bp in length  
\* 149738 149837: gap of unknown length  
\* 149838 162513: contig of 12676 bp in length  
\* 162514 162613: gap of unknown length  
\* 162614 178181: contig of 15568 bp in length.

FEATURES

source  
1. .178181  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="RPG1-11"  
/clone="RP11-415A13"  
misc\_feature  
1. .1401  
/note="assembly\_name:Contig13"  
misc\_feature  
1502. .2671  
/note="assembly\_name:Contig17"  
misc\_feature  
2772. .4499  
/note="assembly\_name:Contig18"  
misc\_feature  
4600. .6309  
/note="assembly\_name:Contig19"  
misc\_feature  
6410. .8601  
/note="assembly\_name:Contig20"  
misc\_feature  
8702. .10415  
/note="assembly\_name:Contig21"  
misc\_feature  
10516. .13064  
/note="assembly\_name:Contig22"  
misc\_feature  
13165. .16156  
/note="assembly\_name:Contig23"  
misc\_feature  
16257. .18638  
/note="assembly\_name:Contig24"  
misc\_feature  
18739. .21462  
/note="assembly\_name:Contig25"  
misc\_feature  
21563. .24282  
/note="assembly\_name:Contig26"  
misc\_feature  
24383. .28368  
/note="assembly\_name:Contig27"  
misc\_feature  
28469. .32249  
/note="assembly\_name:Contig28"  
misc\_feature  
32350. .36555  
/note="assembly\_name:Contig29"  
misc\_feature  
36556. .42852  
/note="assembly\_name:Contig30"  
misc\_feature  
42953. .47243  
/note="assembly\_name:Contig31"

clone\_end:SP6  
vector\_side:right  
47344. .53015  
/note="assembly\_name:Contig33  
clone\_end:T7  
vector\_side:right  
53116. .59822  
/note="assembly\_name:Contig32"  
59923. .65341  
/note="assembly\_name:Contig34"  
65442. .70898  
/note="assembly\_name:Contig35"  
70999. .76921  
/note="assembly\_name:Contig36"  
77022. .85604  
/note="assembly\_name:Contig37"  
85705. .90872  
/note="assembly\_name:Contig38"  
90973. .97315  
/note="assembly\_name:Contig39"  
97416. .105928  
/note="assembly\_name:Contig40"  
106029. .116153  
/note="assembly\_name:Contig41"  
116254. .127251  
/note="assembly\_name:Contig42"  
127352. .136839  
/note="assembly\_name:Contig43"  
136940. .142282  
/note="assembly\_name:Contig44"  
142383. .149737  
/note="assembly\_name:Contig45"  
149838. .162513  
/note="assembly\_name:Contig46"  
162614. .178181  
/note="assembly\_name:Contig47"  
BASE COUNT 52592 a 34925 c 35928 g 51602 t 3134 others  
ORIGIN

Query Match 20.7%; Score 215; DB 77; Length 178181;  
Best Local Similarity 99.6%; Pred. No. 53e-114; Indels 0; Gaps 0;  
Matches 265; Conservative 0; Mismatches 1;

QY 216 gacctggactacttttctatatactgtgtgctaggatcgattacagtagtgactcaag 275  
|||||  
Db 53933 GACCTGGACTACTTTTTCATATACTGTGTGCTAGGATCGATTACAGTAGTGACTCAAG 53992  
|||||  
QY 276 gaaacgtgtgaagtactactgtgtccaaaatcgatccatccccaagtaaccaagatacaac 335  
|||||  
Db 53993 GAAACGTGTGAAGTACTACGTGTCCAAATCGTTCATCCATCCCAAGTACCAAGATACAAC 54052  
|||||  
QY 336 ggcagacrtgccttgttgaactgtctctcaagtccaccttcactctgcatcctgcc 395  
|||||  
Db 54053 GGCAGACGTGCCCTTGTGAAACTGTCTCTCAAGTCACTTCTGCGCATCCTGCC 54112  
|||||  
QY 396 tatttgcttggccagtggtcacaaagcagttggcaattccaccctttgttgggtgaocgg 455  
|||||  
Db 54113 TATTGTGTTGCCAGTGTCAAAAGCAGTTGGCAATTCACCCCTTTTGTGGTGACGG 54172  
|||||  
QY 456 atggggaaaagttaaggaaagttcag 481  
|||||  
Db 54173 ATGGGAAAAGTTAAGGAAAAGTTCAG 54198  
|||||

RESULT 4

AC026556/c  
LOCUS  
DEFINITION  
Homo sapiens chromosome 4 clone RP11-731D1 map 4, WORKING DRAFT  
SEQUENCE, 30 unordered pieces.  
ACCESSION  
AC026556.3 GI:8112965  
VERSION  
HTGS\_PHASE1; HTGS\_DRAFT.



```
misc_feature 9043. .9621
/note="assembly_fragment
clone_end:77
vector_side:right"
9722. .11160
/note="assembly_fragment"
11261. .13550
/note="assembly_fragment"
13651. .17467
/note="assembly_fragment"
17568. .20973
/note="assembly_fragment"
21074. .23243
/note="assembly_fragment"
23344. .26526
/note="assembly_fragment"
26627. .31028
/note="assembly_fragment"
31129. .38352
/note="assembly_fragment"
38453. .43015
/note="assembly_fragment"
43116. .49226
/note="assembly_fragment"
49327. .54582
/note="assembly_fragment"
54683. .60160
/note="assembly_fragment"
60261. .66143
/note="assembly_fragment"
66244. .74231
/note="assembly_fragment"
74332. .80835
/note="assembly_fragment
clone_end:SP6
vector_side:right"
80936. .91055
/note="assembly_fragment"
91156. .101763
/note="assembly_fragment"
101864. .113795
/note="assembly_fragment"
113896. .125368
/note="assembly_fragment"
125469. .145206
/note="assembly_fragment"
145307. .162142
/note="assembly_fragment"

Query Match 5.0%; Score 52; DB 70; Length 180155;
Best Local Similarity 100.0%; Pred. No. 1.7e-18;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgggcctgtgctgtgcttcacgtctctctctctgtggtggtctcag 52
|||||
Db 21490 ATGGCCCTGCTGCTGTGCTTCACGCTCTCTCTCTGCTGGGATCTCAG 21439

RESULT 5
AC004847 AC004847 122280 bp DNA PRI 25-MAR-2001
LOCUS Homo sapiens clone RP4-647J21, complete sequence.
ACCESSION AC004847
VERSION AC004847.3 GI:13446337
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 122280)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished

misc_feature 2 (bases 1 to 122280)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 122280)
Waterston,R.H.
Direct Submission
Submitted (15-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 122280)
Waterston,R.
Direct Submission
Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 122280)
Waterston,R.H.
Direct Submission
Submitted (25-MAR-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 25, 2001 this sequence version replaced gi:9211311.
Center project name: H_DJ0647J21.
FEATURES
Source
1. .122280
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP4-647J21"
BASE COUNT 32248 a 30020 c 28671 g 31341 t
ORIGIN

Query Match 2.0%; Score 21; DB 85; Length 122280;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 ctgctctctctgctggtggtc 48
|||||
Db 114305 CTGCTCTCTCTGCTGGGGATC 114325

RESULT 6
AC015280/c AC015280 159468 bp DNA HTG 16-NOV-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION pieces.
ACCESSION AC015280
VERSION AC015280.1 GI:6436055
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 159468)
Adams,M. and Venter,J.C.
Direct Submission
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10213506 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers
source 1. .159468
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 47335 a 32541 c 32291 g 47301 t
ORIGIN
```

Drosophila melanogaster BAC library, partial ECORI in pBAC3.6)"

BASE COUNT 50404 a 34757 c 34568 g 50553 t  
ORIGIN

Query Match 2.0%; Score 21; DB 63; Length 159468;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 434 caccctttgttggtgacgcg 454  
|||||  
Db 55000 CACCCTTTTGTGGGTGACCG 54980

Query Match 2.0%; Score 21; DB 4; Length 170282;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7  
AC006467/c  
LOCUS  
DEFINITION  
Drosophila melanogaster, chromosome 2L, region 40A-40C, BAC clone  
AC006467  
AC006467.12 GI:13374635  
HTG.  
SOURCE  
ORGANISM  
Drosophila melanogaster  
fruit fly.

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 170282)  
Celnik, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,  
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,  
Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,  
Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,  
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,  
Ferreira, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,  
Gonzalez, M., Houch, J., Hoskins, R.A., Hostin, D., Howland, T.J.,  
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,  
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,  
Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,  
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,  
Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,  
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Unpublished  
Sequencing of Drosophila chromosome 2L, region 40A-40C  
2 (bases 1 to 170282)  
Celnik, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,  
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Sequelra, A., Sethi, H., Shrir, E.,  
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
Rubin, G.M.

Unpublished  
Sequencing of Drosophila chromosome 2L, region 40A-40C  
2 (bases 1 to 170282)  
Celnik, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,  
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Sequelra, A., Sethi, H., Shrir, E.,  
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
Rubin, G.M.

Direct Submission  
Submitted (29-JAN-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Mar 17, 2001 this sequence version replaced gi:6136329.

Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720  
This sequence was assembled using end sequences from a whole genome  
shotgun and from subclones of this BAC and its neighboring clones.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdgpf@fruitfly.berkeley.edu](mailto:bdgpf@fruitfly.berkeley.edu).

FEATURES  
source  
Location/Qualifiers  
1. 170282  
/organism="Drosophila melanogaster"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
/map="40A-40C"  
/clone="BACR03L08 (D532)"  
/clone\_lib="RPCI-98 (Roswell Park Cancer Institute

RESULT 8  
AC084407  
LOCUS  
DEFINITION  
Mus musculus clone RP23-271013 strain C57BL6/J, WORKING DRAFT  
SEQUENCE, 37 unordered pieces.  
AC084407  
AC084407.6 GI:13194209  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 252527)  
Grills, G., Han, J., Montgomery, K.T., Chiu, D., Decker, J., Fusina, M.,  
Goltz, J., Halder, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J.,  
Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.

Unpublished  
High Throughput Mouse Sequencing  
2 (bases 1 to 252527)  
Grills, G., Han, J., Montgomery, K.T., Chiu, D., Decker, J., Fusina, M.,  
Goltz, J., Halder, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J.,  
Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.

Direct Submission  
Submitted (01-NOV-2000) Department of Molecular Genetics, Albert  
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
Bronx, NY 10461, USA  
On Mar 3, 2001 this sequence version replaced gi:11177940.

-----Genome Center  
Center: Albert Einstein College of Medicine  
Center Code: AECOM  
Web site:  
[http://sequence.aecom.yu.edu/cgi-](http://sequence.aecom.yu.edu/cgi-bin/ws.exe/mouseDB/mouseSEQ/mouseseqtable.hts)  
[bin/ws.exe/mouseDB/mouseSEQ/mouseseqtable.hts](http://sequence.aecom.yu.edu/cgi-bin/ws.exe/mouseDB/mouseSEQ/mouseseqtable.hts)  
Contact: [htgs@sequence.aecom.yu.edu](mailto:htgs@sequence.aecom.yu.edu)  
-----Summary Statistics

Center project name: APT  
Sequencing vector: pUC18; L08752  
Chemistry: Dye-terminator Big Dye; 100%  
\*Consensus quality: 240255 at least Q20  
\*Consensus quality: 236058 at least Q30  
\*Consensus quality: 228317 at least Q40  
\*\*Estimated insert size: agarose-FP - N/A  
Quality coverage: 251807 - sum-of-contigs  
Quality coverage: agarose-FP - N/A  
Quality coverage: 7 x sum-of-contigs - N/A  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 37 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 29149: contig of 29149 bp in length  
\* 29150 29169: gap of unknown length  
\* 29170 51849: contig of 22680 bp in length  
\* 51850 51869: gap of unknown length

[illegible]

Query Match 2.08; Score 21; DB 77; Length 252527;

Best Local Similarity 100.0%; Pred. No. 2.5;		MRNA	
Matches	21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy 969 tagtccagggcagagaact 989			
Db 165239 TAGTCCAGGGCAGAGAACT 165259		gene	
RESULT 9			
AE003781/c			
LOCUS	AE003781 311800 bp DNA INV 04-OCT-2000		
DEFINITION	Drosophila melanogaster genomic scaffold 14200001386030 section 1	CDS	
ACCESSION	AE003781 AE002725		
VERSION	AE003781.2 GI:10726338		
KEYWORDS	HTG		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
1 (bases 1 to 311800)			
ADAMS,M.D., Ceiniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazet,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abril,J.F., Agapayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.N., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C., Buesam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieria,S., Fleischmann,W., Foster,C., Gabrielian,A.E., Garg,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwan,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusskern,D.R., Pacleib,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svirkas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A., Weinstein,G.M., Weissbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.		/chromosome="2L" join(16241..16424,17650..17976,18828..19049,19106..>19277) /gene="CG3305" /product="CT11115" /db_xref="FLYBASE:FBan0003305" /db_xref="FLYBASE:FBgn0032949" <16241..>19277 /gene="CG3305" /map="39E1-39E2" /db_xref="FLYBASE:FBan0003305" /db_xref="FLYBASE:FBgn0032949" join(16358..16424,17650..17976,18828..19049,19106..>19277) /gene="CG3305" /note="CG3305 gene product" /codon_start=1 /db_xref="FLYBASE:FBan0003305" /db_xref="FLYBASE:FBgn0032949" /protein_id="AAF57215.1" /db_xref="GI:7302116" /translation="MFANKLLTCSALLMLFLLSTVFSQKLGIPRVKNVSDLFLE QANPLTSSSTISTSTSTTEKPIITPRSTITPRSTSTPSTSPSTAPVISTIAPOP YPQSIGAWNTSCIMLQMAQLNFTYEAREGNTGLNIPSNASVEAECKSQTFQ IHLWGPETSKOSLIMYFNKNDTIVLSFMQIHLALLPEDFDAKENOTVOLITRSDG AFKTPENMSYHCTRQVKINMTETLDAEQILIGHSVSHVQVEAFR" /complement(<20068..>20397) /gene="CG14465" /product="CT34156" /db_xref="FLYBASE:FBan0014465" /db_xref="FLYBASE:FBgn0032950" /evidence=not_experimental complement(join(<20068..20412,21319..21392,21594..22485, 24480..24679,29305..30046,30136..30164,32793..>33192)) /gene="CG12548" /product="CT34155" /db_xref="FLYBASE:FBan0012548" /db_xref="FLYBASE:FBgn0032952" complement(<20068..>20397) /gene="CG14465" /map="39E1-39E2" /db_xref="FLYBASE:FBan0014465" /db_xref="FLYBASE:FBgn0032950" /evidence=not_experimental complement(<20068..>33192) /gene="CG12548" /map="39E1-39E2" /db_xref="FLYBASE:FBan0012548" /db_xref="FLYBASE:FBgn0032952" complement(20068..20397) /gene="CG14465" /note="CG14465 gene product" /codon_start=1 /db_xref="FLYBASE:FBan0014465" /db_xref="FLYBASE:FBgn0032950" /evidence=not_experimental /protein_id="AAF57217.1" /db_xref="GI:7302118" /translation="MASHCVNQLHEIGSIKKNAPGLINAGIVELDSHNLILAREFEG ALQOPMNFARYNGLVALQNDYELAEERFELKQELMPSSGHVFCQQLERRISS KEOQRNN" complement(join(20068..20412,21319..21392,21594..22485, 24480..24679,29305..30046,30136..30164,32793..>33192)) /gene="CG12548" /note="CG12548 gene product" /codon_start=1 /db_xref="FLYBASE:FBan0012548" /db_xref="FLYBASE:FBgn0032952" /protein_id="AAF57216.1" /db_xref="GI:7302117" /translation="WHIEALNTYSIMAKNMFPHVNQLKLMGNIIYSNGIYQAVKM IRMADSVPKLSQLQRKIRENIGILFIRMGSYSDAASSFEFITERANIRSSHLILL CYFALGDVEKVKLRRLCDVQTEAIESDMKEHNTFTSTVFNLAEQERSMHLL ALNTYSIMAKNMFPHVNQLKLMGNIIYSNGIYQAVKMIRMALDSVPKLSQLRLK IRENIGILFIRMGSYSDAASSFEFITERANIRSSHLILLCYFALGDVEKVKLAPRL	
TITLE		CDS	
JOURNAL			
MEDLINE			
REFERENCE			
ADAMS,M.D., Ceiniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.			
Direct Submission			
TITLE			
JOURNAL			
Rockville, MD, USA			
On Oct 9, 2000 this sequence version replaced gi:7302115.			
COMMENT			
LOCATION/Qualifiers			
1..311800			
/organism="Drosophila melanogaster"			
/db_xref="Taxon:7227"			





```

/gene="entB"
/codon_start=1
/transl_table=11
/product="enterocin B prepeptide"
/protein_id="AAC27519.1"
/db_xref="GI:3342122"
/translation="MONVKELSTKEMKQIIGGENDHRMPNELNRPNNLSKGGAKCGAA
IAGGLFGIPKGPGLAWAGLANVSKCN"
complement(1491..1667)
/note="Orf5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC27520.1"
/db_xref="GI:3342123"
/translation="MNLKKNLEYNLCIFLAVIINLGLFIFSETILSKILLIAIVLL
VIPNFMQKKRKN"
complement(1690..1881)
/note="Orf6"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC27521.1"
/db_xref="GI:3342124"
/translation="MFKIIRYSSVFILATCMVPLFIFFEYSMPKILSLIVILLN
VVGILVKELGKMGFCWVD"
complement(1881..2000)
/note="putative induction factor; Orf7"
/codon_start=1
/transl_table=11
/product="putative pheromone"
/protein_id="AAC27522.1"
/db_xref="GI:3342125"
/translation="MQNVKLSMKEIKQTCGGAGKNLIYGMGYLRSCNRL"

BASE COUNT      756 a   392 c   301 g   776 t
ORIGIN

Query Match      1.9%; Score 20; DB 1; Length 2225;
Best Local Similarity 100.0%; Pred.No.7.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  819 cttctctgactctgttc 838
      |||||
DB  546 CTCTCTGACTTCTGTTC 565

RESULT 11
AF121254
LOCUS      AF121254      12018 bp      DNA      BCT      06-MAY-1999
DEFINITION      Enterococcus faecium cysteine aminopeptidase (pepC) gene, partial
                  cds; phospho-beta-glucosidase BglB (bglB), beta-glucoside specific
                  transport protein (bglS), transcription antiterminator (bglR),
                  enterocin B precursor (entB), enterocin B immunity protein (entI),
                  and induction factor-like protein genes, complete cds; and unknown
                  genes.
ACCESSION      AF121254
VERSION        AF121254.1  GI:4704705
KEYWORDS
SOURCE
ORGANISM      Enterococcus faecium.
                  Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
                  Enterococcus.
REFERENCE
AUTHORS      Franz,C.M.A.P., Worobo,R.W., Quadri,L.E.N., Schillinger,U.,
                  Holzapfel,W.H., Vederas,J.C. and Stiles,M.E.
TITLE        Atypical genetic locus associated with constitutive production of
                  enterocin B by Enterococcus faecium BFE 900
JOURNAL      Appl. Environ. Microbiol. 65 (5), 2170-2178 (1999)
MEDLINE      9924046
REFERENCE
AUTHORS      Franz,C.M.A.P., Worobo,R.W., Quadri,L.E.N., Schillinger,U.,
                  Holzapfel,W.H., Vederas,J.C. and Stiles,M.E.
```

```

TITLE      Direct Submission
JOURNAL    Submitted (19-JAN-1999) Department of Agricultural, Food and
                  Nutritional Science, University of Alberta, 4-10 Agriculture
                  Forestry Center, Edmonton, Alberta T6G 2P5, Canada
FEATURES
source
1..12018
/organism="Enterococcus faecium"
/strain="BFE 900"
/db_xref="taxon:1352"
complement(<1..675)
/gene="pepC"
complement(<1..675)
/gene="pepC"
/note="PepC"
/codon_start=1
/transl_table=11
/product="cysteine aminopeptidase"
/protein_id="AAD28219.1"
/db_xref="GI:4704706"
/translation="MKAITPDMTTRKFKNQFQANIKQOALQRAVVKNGITASAENVSAH
VENTPVFSIDLATGKVNQMSGRCMFAALNTRHKMLNOFHLKEFELSONYTFWD
KYEKANYEYENIATGKEAIDSRKVAFLLATPODGGQWDMLVSLFQKGVVPKSVMP
ESANSNSRDLNNYLKMLKRDVAVLRKVAOGETLDATEERKEEMLESIIYNFLAISL
GTPPKPEF"
354..494
/note="orf17"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD28220.1"
/db_xref="GI:4704707"
/translation="MPFHIYPRKCNFVIQTLNGIDSASVDGMCVSQOTSSNAPIA
FD"
859..1134
/note="orf16"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD28221.1"
/db_xref="GI:4704708"
/translation="MACIIRKPRRTSYNKIAFKILIERKYYNVRKKNMKNLIAG
GFLFVGGIGLMTDNFNVIPFLGAACALIGVMLIYFVSDDGKG"
complement(1098..1259)
/note="orf15"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD28222.1"
/db_xref="GI:4704709"
/translation="MQRKRNPITWYSFSFIIMNQAFCHVCCNLRYERVELATHNQ
PVPSSLNPK"
1131..1283
/note="orf14"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD28223.1"
/db_xref="GI:4704710"
/translation="MIMSQLYSFVKITTYVTKTESLIHNNKRNKRNPSDWISFSLHK
LYLLRQ"
1765..2394
/note="orf13"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD28224.1"
/db_xref="GI:4704711"
/translation="MKIVFLHGLGODEQSKVDIVSPNSYCKSLSLFNHLNSDTDA
TLDDLTKEIVOLNOIKEPFILCGLSLGAVALNYLATPSVLYKKVLAPOFESPNR
LLLTQNIYFYCVPSKIEKELGSLSKQDILTLSLETLDCKQLTSRLPLILCGTK
DHANKPAARLYKLTSTNGRLLFIRKNGKHELNKPEIMAQAQRMSFIENS"
2414..2596
/note="orf12"
CDS
```



```
/strain="129/SvJ"
/db_xref="taxon:10090"
/chromosome="8"
/map="28cm"
7414..15443
/feature="Ucp"
7414..7644
/feature="Ucp"
7414..7770
/feature="Ucp"
/number=1
join(7645..7770,8430..8628,10920..11120,11215..11316,
12185..12365,14867..14981)
/feature="Ucp"
/codon_start=1
/product="uncoupling protein"
/protein_id="AAB07367.1"
/db_xref="GI:1519065"
/translation="MVNPTTSEVQPTMGKIFSGVSAACLAADIITFPDLDTAKVRLIQ
GEGQASSTIRYKGLVIGTITTLAKTEGLKPLXGLPAGIQRQISFASLRIGLYDSVOEY
FSSGRETASLGNKISAGLMTGGVAVFIGOPTVEVVKVQMOASHLHGKIPRYGTGYN
YRVVATTESLTLKGTTPNLMRNVIINCTELVTDLMKGLYNNKILADDPCHLLS
ALVAGFCITLASPDVVKTRFINSPLQOYPSVPSCAMSTTKEGTFAFFKGFVASFL
RLGSWNVIMFVCFEQLKKELMKSRQTVDCIT"
8430..8628
/feature="Ucp"
/number=2
10920..11120
/feature="Ucp"
/number=3
11215..11316
/feature="Ucp"
/number=4
12185..12365
/feature="Ucp"
/number=5
14867..15059
/feature="Ucp"
/number=6
15038
/feature="Ucp"
15060
/feature="Ucp"
15443
/feature="Ucp"
BASE COUNT 4538 a 3679 c 3512 g 4369 t
ORIGIN

Query Match 1.9%; Score 20; DB 94; Length 16098;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 tgtcacaaagcagttgcaa 430
|||||
Db 5665 TGTCACAAAGCAGTTGGCAA 5646

RESULT 13
AC006957
LOCUS AC006957 Homo sapiens chromosome 5q, Bac clone 104n10 (LBNL H295), complete
sequence.
DEFINITION AC006957.1 GI:4337159
ACCESSION AC006957.1
VERSION AC006957.1
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 66494)
AUTHORS Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Kadner,K.,
Miguel,T., Miller,C., Pitluck,S., Pollard,M., Subramanian,S.,
```

```
Wheeland,A. and Martin,C.H.
Sequencing of human chromosome 5
Unpublished
2 (bases 1 to 66494)
Ricke,D.O.
Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
Unpublished
3 (bases 1 to 66494)
Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Kadner,K.,
Miguel,T., Miller,C., Pitluck,S., Pollard,M., Subramanian,S.,
Wheeland,A. and Martin,C.H.
Direct Submission
Submitted (05-MAR-1999) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
Sequence submitted by:
DOE Joint Genome Institute.
Location/Qualifiers
1..66494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Bac 104n10"
complement(2958..3087)
/rpt_family="MER5"
2958..3086
/rpt_family="MER5"
3412..4473
/rpt_family="MER11A"
8053..8184
/rpt_family="Alu"
complement(11515..17519)
/rpt_family="L1"
complement(16075..16293)
/rpt_family="MER25"
19037..19398
/rpt_family="MLT1"
complement(20322..20466)
/rpt_family="MER5"
complement(21945..22351)
/rpt_family="MLT1"
22831..22974
/feature="GRAIL 2 excellent exon, frame 1"
complement(26378..26483)
/rpt_family="MIR"
26626..26787
/rpt_family="MER20"
complement(29874..30129)
/rpt_family="Alu"
complement(30570..30725)
/rpt_family="MIR"
33681..33806
/rpt_family="MIR"
35213..35318
/rpt_family="MIR"
35685..35872
/rpt_family="L1"
36136..36218
/rpt_family="MLT1"
36155..36234
/rpt_family="THE1"
36239..36312
/rpt_family="MSTAR"
37523..37815
/rpt_family="Alu"
39488..41654
/rpt_family="L1"
42957..43133
/rpt_family="MIR"
complement(45673..45765)
/rpt_family="MIR"
complement(50759..51019)
/rpt_family="Alu"
```

misc_feature		53409..53458		/note="GRAIL 2 excellent exon, frame 1"	
repeat_region		62104..62171		/rpt_family="MIR"	
repeat_region		complement(62348..66465)		/rpt_family="LI"	
BASE COUNT		17680 a 13908 c 14466 g 20440 t			
ORIGIN					
Query Match		1.9%; Score 20; DB 86; Length 66494;			
Best Local Similarity		100.0%; Pred. No. 8.9;			
Matches 20; Conservative		0; Mismatches 0; Indels 0; Gaps 0;			
QY 619		ggtgatactactaaaacatgaa 638			
Db 47747		GGTGATACTACAAACATGAA 47766			
RESULT 14					
AC018760					
LOCUS		AC018760 95855 bp DNA PRI 18-APR-2000			
DEFINITION		Homo sapiens chromosome 5 clone CTB-99P17, complete sequence.			
ACCESSION		AC018760			
VERSION		AC018760.4 GI:7019613			
KEYWORDS		HTG.			
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS		DOE Joint Genome Institute and Stanford Human Genome Center.			
TITLE		Direct Submission			
JOURNAL		Unpublished			
REFERENCE		2 (bases 1 to 95855)			
AUTHORS		DOE Joint Genome Institute.			
TITLE		Direct Submission			
JOURNAL		Submitted (19-DEC-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
REFERENCE		3 (bases 1 to 95855)			
AUTHORS		DOE Joint Genome Institute and Stanford Human Genome Center.			
TITLE		Direct Submission			
JOURNAL		Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
COMMENT		On Feb 22, 2000 this sequence version replaced g1:6957626. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov			
Finishing Completed at Stanford Human Genome Center		www.shgc.stanford.edu			
Quality: Phrap Quality >=40 99.4% of Sequence;		Estimated Total Number of Errors is 0.6.			
FEATURES		Location/Qualifiers			
source		1..95855			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
		/chromosome="5"			
		/clone="CTB-99P17"			
BASE COUNT		27700 a 22689 c 21250 g 24216 t			
ORIGIN					
Query Match		1.9%; Score 20; DB 87; Length 95855;			
Best Local Similarity		100.0%; Pred. No. 9;			
Matches 20; Conservative		0; Mismatches 0; Indels 0; Gaps 0;			
QY 401		gcttgcacagtgtcacaaag 420			
Db 85622		GCTTGCACAGTGTCAAAAG 85641			

RESULT 15	AC010866	LOCUS	AC010866 112622 bp DNA HTG 03-SEP-2000
DEFINITION	Homo sapiens chromosome 15 clone RP11-194NT, *** SEQUENCING IN PROGRESS ***, 82 unordered pieces.		
ACCESSION	AC010866	VERSION	AC010866.2 GI:9965523
KEYWORDS	HTG: HTGS_PHASE1.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 112622)		
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A., Nesbitt,R., Traicoff,R. and Hood,L.	Sequencing of human chromosome 15 D15S114-D15S115 region Unpublished		
TITLE	2 (bases 1 to 112622)		
JOURNAL	Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A., Ratcliffe,A., Shaffer,T. and Hood,L.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (25-SEP-1999) Multimegabase Sequencing Center, University of Washington, PO BOX 357730, Seattle, WA 98195, USA		
TITLE	On Sep 3, 2000 this sequence version replaced gi:5923652.		
JOURNAL	----- Genome Center		
COMMENT	Center: Multimegabase Sequencing Center		
	Center code: UMWSC		
	Web site: http://chroma.mbt.washington.edu/msg_www		
	Contact: leerowen@systemsbiology.org		
	----- Summary Statistics		
	Sequencing vector: pUC18; L08752		
	Chemistry: Dye-terminator Big Dye; 90% of reads		
	Chemistry: Dye-terminator Big Dye; 10% of reads		
	Assembly program: Phrap; version 0.990399		
	-----		
	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 82 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		
	* be preserved.		
	* 1 1119: contig of 1119 bp in length		
	* 1120 1219: gap of unknown length		
	* 1220 2227: contig of 1008 bp in length		
	* 2328 2327: gap of unknown length		
	* 3128 3127: contig of 800 bp in length		
	* 3228 3227: gap of unknown length		
	* 4071 4071: contig of 844 bp in length		
	* 4072 4071: gap of unknown length		
	* 4172 4068: contig of 697 bp in length		
	* 4869 4968: gap of unknown length		
	* 5783 5782: contig of 814 bp in length		
	* 5883 5882: gap of unknown length		
	* 6599 6798: gap of unknown length		
	* 6799 7346: contig of 548 bp in length		
	* 7347 7446: gap of unknown length		
	* 7447 8428: contig of 982 bp in length		
	* 8429 8528: gap of unknown length		
	* 8529 9241: contig of 713 bp in length		
	* 9242 9341: gap of unknown length		
	* 9342 10133: contig of 792 bp in length		
	* 10134 10233: gap of unknown length		
	* 10234 11097: contig of 864 bp in length		
	* 11098 11197: gap of unknown length		
	* 11198 12369: contig of 1172 bp in length		
	* 12370 12469: gap of unknown length		

RESULT 15

AC010866

LOCUS

DEFINITION

PROGRESS \*\*\*

AC010866

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AC010866 112622 bp DNA HTG 03-SEP-2000  
Homo sapiens chromosome 15 clone RP11-194N7, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*

AC010866 2 GI:9965523  
HTG: HTGS\_PHASE1.  
human.

Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 112622)  
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,  
Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A.,  
Nesbitt,R., Traicoff,R. and Hood,L.

Sequencing of human chromosome 15 D15S114-D15S115 region  
Unpublished

2 (bases 1 to 112622)  
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,  
Bloom,S., Dors,M., Dickhoff,R., Harrison,G., James,R., Lasky,S.,  
Madan,A., Ratcliffe,A., Shaffer,T. and Hood,L.

Direct Submission  
Submitted (25-SEP-1999) Multimegabase Sequencing Center, University  
of Washington, PO BOX 357730, Seattle, WA 98195, USA

On Sep 3, 2000 this sequence version replaced gi:5923652.  
----- Genome Center

Center: Multimegabase Sequencing Center  
Center code: UWMSC  
Web site: http://chroma.mbt.washington.edu/msg\_www

Contact: leerowen@systemsbiology.org  
----- Summary Statistics

Sequencing vector: pUC18: L08752  
Chemistry: Dye-terminator Big Dye; 90% of reads  
Chemistry: Dye-primer Big Dye; 10% of reads

Assembly program: Phrap; version 0.990399  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 82 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1119: contig of 1119 bp in length  
\* 1120: gap of unknown length  
\* 1220: contig of 1008 bp in length  
\* 2328: gap of unknown length  
\* 2328: contig of 800 bp in length  
\* 3128: gap of unknown length  
\* 3228: contig of 844 bp in length  
\* 4072: gap of unknown length  
\* 4172: contig of 697 bp in length  
\* 4868: gap of unknown length  
\* 4969: contig of 814 bp in length  
\* 5783: gap of unknown length  
\* 5882: contig of 816 bp in length  
\* 6698: gap of unknown length  
\* 6799: contig of 548 bp in length  
\* 7347: gap of unknown length  
\* 7447: contig of 982 bp in length  
\* 8429: gap of unknown length  
\* 8529: contig of 713 bp in length  
\* 9242: gap of unknown length  
\* 9342: contig of 792 bp in length  
\* 10134: gap of unknown length  
\* 10234: contig of 864 bp in length  
\* 11098: gap of unknown length  
\* 11198: contig of 1172 bp in length  
\* 12369: contig of 1172 bp in length  
\* 12469: gap of unknown length

```
* 12470 13731: contig of 1262 bp in length
* 13732 13831: gap of unknown length
* 13832 14659: contig of 828 bp in length
* 14660 14759: gap of unknown length
* 14760 15788: contig of 1029 bp in length
* 15789 15888: gap of unknown length
* 15889 16725: contig of 837 bp in length
* 16726 16825: gap of unknown length
* 16826 17690: contig of 865 bp in length
* 17691 17790: gap of unknown length
* 17791 18698: contig of 908 bp in length
* 18699 18798: gap of unknown length
* 18799 19524: contig of 726 bp in length
* 19525 19624: gap of unknown length
* 19625 20307: contig of 683 bp in length
* 20308 20407: gap of unknown length
* 20408 21157: contig of 750 bp in length
* 21158 21257: gap of unknown length
* 21258 21969: contig of 712 bp in length
* 21970 22069: gap of unknown length
* 22070 22796: contig of 727 bp in length
* 22797 22896: gap of unknown length
* 22897 23658: contig of 762 bp in length
* 23659 23758: gap of unknown length
* 23759 24617: contig of 859 bp in length
* 24618 24717: gap of unknown length
* 24718 25481: contig of 764 bp in length
* 25482 25581: gap of unknown length
* 25582 26338: contig of 757 bp in length
* 26339 26438: gap of unknown length
* 26439 27207: contig of 769 bp in length
* 27208 27307: gap of unknown length
* 27308 28401: contig of 1094 bp in length
* 28402 28501: gap of unknown length
* 28502 28593: contig of 1092 bp in length
* 28594 29694: gap of unknown length
* 29694 30844: contig of 1151 bp in length
* 30845 30944: gap of unknown length
* 30945 31651: contig of 707 bp in length
* 31652 31751: gap of unknown length
* 31752 32454: contig of 703 bp in length
* 32454 32554: gap of unknown length
* 32555 33886: contig of 1332 bp in length
* 33887 33986: gap of unknown length
* 33987 35680: contig of 1694 bp in length
* 35681 35780: gap of unknown length
* 35781 36627: contig of 847 bp in length
* 36628 36727: gap of unknown length
* 36728 38304: contig of 1577 bp in length
* 38305 38404: gap of unknown length
* 38405 39697: contig of 1293 bp in length
* 39698 39797: gap of unknown length
* 39798 40885: contig of 1088 bp in length
* 40886 40985: gap of unknown length
* 40986 42572: contig of 1587 bp in length
* 42573 42672: gap of unknown length
* 42673 43937: contig of 1265 bp in length
* 43938 44037: gap of unknown length
* 44038 45317: contig of 1280 bp in length
* 45318 45417: gap of unknown length
* 45418 46586: contig of 1169 bp in length
* 46587 46686: gap of unknown length
* 46687 47528: contig of 842 bp in length
* 47529 47628: gap of unknown length
* 47629 48692: contig of 1064 bp in length
* 48693 48792: gap of unknown length
* 48793 49852: contig of 1060 bp in length
* 49853 49952: gap of unknown length
* 49953 51146: contig of 1194 bp in length
* 51147 52246: gap of unknown length
* 52237 52337: contig of 991 bp in length
* 52338 52395: contig of 958 bp in length
```

```
* 53296 53395: gap of unknown length
* 53396 54810: contig of 1415 bp in length
* 54811 54910: gap of unknown length
* 54911 56108: contig of 1198 bp in length
* 56109 56208: gap of unknown length
* 56209 57541: contig of 1333 bp in length
* 57542 57641: gap of unknown length
* 57642 59153: contig of 1512 bp in length
* 59154 59253: gap of unknown length
* 59254 60345: contig of 1092 bp in length
* 60346 60445: gap of unknown length
* 60446 62262: contig of 1817 bp in length
* 62263 62362: gap of unknown length
* 62363 63650: contig of 1288 bp in length
* 63651 63750: gap of unknown length
* 63751 65099: contig of 1349 bp in length
* 65100 65199: gap of unknown length
* 65200 66669: contig of 1470 bp in length
* 66670 66769: gap of unknown length
* 66770 67835: contig of 1066 bp in length
* 67836 67935: gap of unknown length
* 67936 68636: contig of 701 bp in length
* 68637 68736: gap of unknown length
* 68737 70126: contig of 1390 bp in length
* 70127 70226: gap of unknown length
* 70227 71496: contig of 1270 bp in length
* 71497 71596: gap of unknown length
* 71597 73367: contig of 1771 bp in length
* 73368 73467: gap of unknown length
* 73468 75753: contig of 2286 bp in length
* 75754 75853: gap of unknown length
* 75854 77323: contig of 1470 bp in length
* 77324 77423: gap of unknown length
* 77424 79098: contig of 1675 bp in length
* 79099 79198: gap of unknown length
* 79199 80173: contig of 975 bp in length
* 80174 80273: gap of unknown length
* 80274 81612: contig of 1339 bp in length
* 81613 81712: gap of unknown length
* 81713 83978: contig of 2266 bp in length
* 83979 84078: gap of unknown length
* 84079 85603: contig of 1525 bp in length
* 85604 85703: gap of unknown length
* 85704 87353: contig of 1650 bp in length
* 87354 87453: gap of unknown length
* 87454 89700: contig of 2247 bp in length
* 89701 89800: gap of unknown length
* 89801 91123: contig of 1323 bp in length
* 91124 91223: gap of unknown length
* 91224 93789: contig of 2566 bp in length
* 93790 93889: gap of unknown length
* 93890 97332: contig of 3443 bp in length
* 97333 97432: gap of unknown length
* 97433 99620: contig of 2188 bp in length
* 99621 99720: gap of unknown length
* 99721 101171: contig of 1451 bp in length
* 101172 101271: gap of unknown length
* 101272 103124: contig of 1853 bp in length
```

Query Match 1.9%; Score 20; DB 61; Length 112622;  
Best Local Similarity 100.0%; Pred. No. 9.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 625 actcaaacatgaagatag 644  
|||||

Db 8104 ACTCAAAACATGAGGATAG 8123

RESULT 16  
AL137009/c  
LOCUS AL137009 113254 bp DNA PRI 18-AUG-2000  
DEFINITION Human DNA sequence from clone RP3-37005 on chromosome 6 Contains  
STSs, GSSs and a CpG island, complete sequence.

ACCESSION	AL137009	repeat_region	6624. .7129
VERSION	AL137009.8	repeat_region	/note="L1M1 repeat: matches 5769. .6297 of consensus"
KEYWORDS	HTG; Cpg island.	repeat_region	7780. .8167
SOURCE	human.	repeat_region	/note="L2 repeat: matches 1915. .2301 of consensus"
ORGANISM	Homo sapiens	repeat_region	8398. .8500
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;	repeat_region	/note="L2 repeat: matches 2100. .2201 of consensus"
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	repeat_region	9745. .10092
TITLE	1 (bases 1 to 113254)	repeat_region	/note="MLT1A1 repeat: matches 1. .365 of consensus"
JOURNAL	Bates,K.	repeat_region	10335. .11116
COMMENT	Direct Submission	repeat_region	/note="L1MB6 repeat: matches 5395. .6175 of consensus"
	Submitted (11-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,	repeat_region	11789. .11857
	CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk	repeat_region	/note="L2 repeat: matches 1745. .1818 of consensus"
	requests: clonerequest@sanger.ac.uk	repeat_region	12452. .12565
	On Jul 28, 2000 this sequence version replaced gi:9407728.	repeat_region	/note="MER85 repeat: matches 1. .138 of consensus"
	During sequence assembly data is compared from overlapping clones.	repeat_region	12924. .13231
	Where differences are found these are annotated as variations	repeat_region	/note="AluJb repeat: matches 1. .304 of consensus"
	together with a note of the overlapping clone name. Note that the	repeat_region	13347. .13438
	variation annotation may not be found in the sequence submission	repeat_region	/note="H14 repeat: matches 1. .93 of consensus"
	corresponding to the overlapping clone, as we submit sequences with	repeat_region	13577. .13642
	only a small overlap as described above.	repeat_region	/note="MLT1H repeat: matches 51. .123 of consensus"
	This sequence has been finished according to sequence map criteria	repeat_region	13643. .13969
	as follows. An attempt is made to resolve all sequencing problems,	repeat_region	/note="AluSx repeat: matches 1. .308 of consensus"
	such as compressions and repeats, but not necessarily within known	repeat_region	13970. .14148
	annotated human repeat sequence elements (e.g. Alu). Where the	repeat_region	/note="MLT1H repeat: matches 123. .268 of consensus"
	sequence is ambiguous, there is an annotation using the 'unsure'	repeat_region	14409. .14462
	feature key.	repeat_region	/note="27 copies 2 mer tt 75% conserved"
	The following abbreviations are used to associate primary accession	repeat_region	14485. .14746
	numbers given in the feature table with their source databases:	repeat_region	/note="AluJb repeat: matches 1. .279 of consensus"
	Em:, EMBL; SW:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information	repeat_region	16057. .16459
	on the WORMPEP database can be found at	repeat_region	/note="MSTA repeat: matches 3. .426 of consensus"
	http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence	repeat_region	16815. .17091
	was generated from part of bacterial clone contigs of human	repeat_region	/note="AluJo repeat: matches 54. .304 of consensus"
	chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping	repeat_region	17232. .17290
	Group. Further information can be found at	repeat_region	/note="L1MA5A repeat: matches 5981. .6038 of consensus"
	http://www.sanger.ac.uk/HGP/Chr6	repeat_region	17291. .17382
	RP3-37005 is from the library RPCI-3 constructed at the Roswell	repeat_region	/note="L1MA5A repeat: matches 6199. .6294 of consensus"
	Park Cancer Institute by the group of Pieter de Jong. For further	repeat_region	17385. .17497
	details see http://bacpac.med.buffalo.edu/	repeat_region	/note="AluJ/monomer repeat: matches 1. .49 of consensus"
	VECTOR: pCYPAC2	repeat_region	17598. .17754
FEATURES	Location/Qualifiers	repeat_region	/note="MIR repeat: matches 64. .230 of consensus"
source	1. .113254	repeat_region	17851. .18147
	/organism="Homo sapiens"	repeat_region	/note="AluJb repeat: matches 3. .303 of consensus"
	/db_xref="taxon:9606"	misc_feature	complement(18027. .18336)
	/chromosome="6"	repeat_region	/note="match: GSS: Em:AQ217130"
	/clone="RP3-37005"	repeat_region	19015. .19137
	/clone_lib="RPCI-3"	repeat_region	/note="MIR repeat: matches 45. .189 of consensus"
	59. .142	repeat_region	19886. .20192
repeat_region	/note="42 copies 2 mer aa 63% conserved"	repeat_region	/note="AluJo repeat: matches 1. .300 of consensus"
repeat_region	1694. .1948	repeat_region	20343. .20478
misc_feature	/note="AluSx repeat: matches 44. .297 of consensus"	repeat_region	/note="L2 repeat: matches 1641. .1780 of consensus"
repeat_region	1695. .1948	repeat_region	20510. .20989
repeat_region	/note="match: GSS: Em:AQ839834"	repeat_region	/note="MLT1F repeat: matches 1. .540 of consensus"
repeat_region	2315. .2991	repeat_region	21739. .21987
repeat_region	/note="L1MA2 repeat: matches 5110. .5787 of consensus"	repeat_region	/note="MLT1D repeat: matches 9. .355 of consensus"
repeat_region	2987. .4128	repeat_region	22889. .22973
repeat_region	/note="L1PA13 repeat: matches 3281. .4421 of consensus"	repeat_region	/note="MER82 repeat: matches 570. .653 of consensus"
repeat_region	4149. .4276	repeat_region	22978. .23169
repeat_region	/note="L1MA2 repeat: matches 5913. .6041 of consensus"	repeat_region	/note="MER82 repeat: matches 83. .277 of consensus"
repeat_region	4274. .4633	repeat_region	23293. .23465
repeat_region	/note="L1PA13 repeat: matches 4439. .4799 of consensus"	repeat_region	/note="FRAM repeat: matches 1. .173 of consensus"
repeat_region	4634. .4942	repeat_region	23315. .23815
repeat_region	/note="AluSx repeat: matches 2. .310 of consensus"	repeat_region	/note="AluSx repeat: matches 1. .300 of consensus"
repeat_region	4943. .5991	repeat_region	23912. .24052
repeat_region	/note="L1PA13 repeat: matches 4799. .5828 of consensus"	repeat_region	/note="MIR repeat: matches 53. .197 of consensus"
repeat_region	5992. .6294	repeat_region	24519. .24598
repeat_region	/note="AluYb8 repeat: matches 1. .304 of consensus"	repeat_region	/note="MADE1 repeat: matches 1. .80 of consensus"
repeat_region	6295. .6622	repeat_region	27328. .27628
	/note="L1PA13 repeat: matches 5828. .6156 of consensus"	repeat_region	/note="AluSx repeat: matches 1. .305 of consensus"
		repeat_region	30329. .30403
		repeat_region	/note="L2 repeat: matches 2607. .2685 of consensus"
		repeat_region	30874. .31022
		repeat_region	/note="MIR repeat: matches 18. .175 of consensus"
		repeat_region	31521. .31720

```
/note="MER20 repeat: matches 27. .215 of consensus"
complement(33394. .33861)
/note="match: STS: Em:G31570"
33394. 34525
/note="CpG island"
/evidence=not_experimental
36033. .36260
/note="MER46A repeat: matches 7. .236 of consensus"
36277. .36589
/note="MSTA repeat: matches 97. .426 of consensus"
36598. .36770
/note="THE1B repeat: matches 94. .259 of consensus"
36771. .37360
/note="LTR1 repeat: matches 203. .785 of consensus"
37393. .37617
/note="LTR1 repeat: matches 32. .258 of consensus"
complement(37659. .37768)
/note="match: GSS: Em:AF011209"
37675. .37772
/note="match: GSS: Em:AQ130909"
37675. .37768
/note="match: GSS: Em:AF011189"
complement(37675. .37772)
/note="match: GSS: Em:AQ531062 Em:AQ533711"
complement(37679. .37768)
/note="match: GSS: Em:AQ314075"
37723. .38206
/note="match: GSS: Em:AQ078123"
37769. .39375
/note="THE1C-internal repeat: matches 3. .1580 of
consensus"
39378. .39755
/note="THE1C repeat: matches 1. .371 of consensus"
39760. .41138
/note="MSTA-internal repeat: matches 1. .1442 of consensus"
41139. .41534
/note="MSTA repeat: matches 1. .426 of consensus"
41686. .42422
/note="L2 repeat: matches 1953. .2705 of consensus"
42540. .42604
/note="MER69 repeat: matches 2436. .2509 of consensus"
42669. .43743
/note="L1MB3 repeat: matches 5103. .6185 of consensus"
44518. .44661
/note="AluJo repeat: matches 170. .303 of consensus"
44662. .44930
/note="AluX repeat: matches 29. .309 of consensus"
44931. .45069
/note="AluJo repeat: matches 1. .170 of consensus"
45591. .45739
/note="LTR16C repeat: matches 155. .330 of consensus"

Query Match 1.9%; Score 20; DB 89; Length 113254;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 agtactacgtgtccaaatc 306
|||||
Db 11827 AGTACTACGTGTCCAAATC 11808

RESULT 17
HSJ193N13/C
LOCUS HSJ193N13 122961 bp DNA PRI 18-FEB-2000
DEFINITION Human DNA sequence from clone RPL-193N13 on chromosome 6q21-22.31
Contains the gene for MAN1A (mannosidase, alpha,
class 1A, member 1), ESTs, STSS, GSSs and a CpG Island, complete
sequence.
ACCESSION AL078600
VERSION AL078600.15 GI:6273536
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
```

```
REFERENCE
AUTHORS Dunn, M.
JOURNAL Direct Submission
COMMENT Submitted (18-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Nov 7, 1999 this sequence version replaced gi:6165363.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep/RPL-193N13 is
from the library RPCI-1 constructed at the Roswell Park Cancer
Institute by the group of Pieter de Jong. For further details see
http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RPL-193N13 This
sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed by the Sanger Centre Chromosome 6
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6.
FEATURES
source
1. .122961
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q21-22.31"
/clone="RPL-193N13"
/clone_lib="RPCI-1"
725. .921
repeat_region
/note="MIR repeat: matches 20. .232 of consensus"
940. .1160
repeat_region
/note="AluJ repeat: matches 1. .110 of consensus"
complement(1161. .1567)
/misc_feature
/note="match: STS: Em:Z77891"
1804. .2104
repeat_region
/note="AluY repeat: matches 1. .301 of consensus"
3455. .3762
repeat_region
/note="AluSg repeat: matches 1. .308 of consensus"
4765. .4948
repeat_region
/note="MIR repeat: matches 1. .199 of consensus"
complement(join(<5197. .5312,10105. .10201,51672. .>52550))
/gene="MAN1A"
/note="match: CDNAS: Em:U04301 Em:U03458
Em:U03457 Em:Y12503 Em:X74837 Em:AF027156; match: ESTs:
Em:A1256535"
/evidence=not_experimental
/product="dJ193N13.1 (mannosidase, alpha, class 1A, member
1)"
gene
complement(5197. .52550)
/gene="MAN1A"
CDS
complement(join(<5197. .5312,10105. .10201,51672. .52274))
/gene="MAN1A"
/note="match: proteins: Sw:P39098 Tr:O60476 Sw:P45700
Tr:O02773 Sw:P33908 Sw:P45701 Sw:P53624 Wp:CE08947
Sw:P53625"
/codon_start=1
/evidence=not_experimental
/product="dJ193N13.1 (mannosidase, alpha, class 1A, member
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 122961)

Dunn, M.  
Direct Submission  
Submitted (18-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk

On Nov 7, 1999 this sequence version replaced gi:6165363.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.

The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/projects/C.elegans/wormpep/RPL-193N13 is  
from the library RPCI-1 constructed at the Roswell Park Cancer  
Institute by the group of Pieter de Jong. For further details see  
http://bacpac.med.buffalo.edu/  
VECTOR: pCYPAC2

This sequence is the entire insert of clone RPL-193N13 This  
sequence was generated from part of bacterial clone contigs of  
human chromosome 6, constructed by the Sanger Centre Chromosome 6  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6.

FEATURES  
source  
1. .122961  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="q21-22.31"  
/clone="RPL-193N13"  
/clone\_lib="RPCI-1"  
725. .921  
repeat\_region  
/note="MIR repeat: matches 20. .232 of consensus"  
940. .1160  
repeat\_region  
/note="AluJ repeat: matches 1. .110 of consensus"  
complement(1161. .1567)  
/misc\_feature  
/note="match: STS: Em:Z77891"  
1804. .2104  
repeat\_region  
/note="AluY repeat: matches 1. .301 of consensus"  
3455. .3762  
repeat\_region  
/note="AluSg repeat: matches 1. .308 of consensus"  
4765. .4948  
repeat\_region  
/note="MIR repeat: matches 1. .199 of consensus"  
complement(join(<5197. .5312,10105. .10201,51672. .>52550))  
/gene="MAN1A"  
/note="match: CDNAS: Em:U04301 Em:U03458  
Em:U03457 Em:Y12503 Em:X74837 Em:AF027156; match: ESTs:  
Em:A1256535"  
/evidence=not\_experimental  
/product="dJ193N13.1 (mannosidase, alpha, class 1A, member  
1)"  
gene  
complement(5197. .52550)  
/gene="MAN1A"  
CDS  
complement(join(<5197. .5312,10105. .10201,51672. .52274))  
/gene="MAN1A"  
/note="match: proteins: Sw:P39098 Tr:O60476 Sw:P45700  
Tr:O02773 Sw:P33908 Sw:P45701 Sw:P53624 Wp:CE08947  
Sw:P53625"  
/codon\_start=1  
/evidence=not\_experimental  
/product="dJ193N13.1 (mannosidase, alpha, class 1A, member





JOURNAL Submitted (06-JUN-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 3 (bases 1 to 123397)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (12-JUN-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 4 (bases 1 to 123397)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 5 (bases 1 to 123397)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT ----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
-----  
Center project name: H\_DJ0587D13  
-----  
NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.  
  
This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.  
  
MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send [mailto:regreen@nhgri.nih.gov](mailto:mailto:regreen@nhgri.nih.gov), or see <http://genome.wustl.edu/gsc>  
  
SOURCE INFORMATION:  
This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., Nature Genetics, 6:84-9 (1994). The library is from one male donor. For further details, see <http://baopac.med.buffalo.edu/>  
The clone is available from Genome Systems, Inc. (<http://www.genomesystems.com>).  
VECTOR: pCYPAC2  
NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP5-118517. Actual start of this clone is at base position 1 of RP4-587D13; actual end is at 123397 of RP4-587D13.  
  
FEATURES  
source  
1..123397  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q11.23-q21.1"  
/clone="RP4-587D13"  
/clone.lib="RPCI-4"  
891..1175  
/rpt\_family="Alu"  
repeat\_region  
1176..1449  
/rpt\_family="Alu"  
complement(28045..28370)  
/evidence=not\_experimental

repeat\_region 1456..1853  
/rpt\_family="MaLR"  
repeat\_region 3415..3714  
/rpt\_family="Alu"  
repeat\_region 3949..4232  
/rpt\_family="Alu"  
repeat\_region 4302..4341  
/rpt\_family="Alu"  
repeat\_region 4348..4420  
/rpt\_family="L1"  
repeat\_region 5226..5576  
/rpt\_family="L1"  
repeat\_region 5577..6112  
/rpt\_family="L1"  
repeat\_region 6178..6530  
/rpt\_family="L1"  
repeat\_region 6534..6804  
/rpt\_family="Alu"  
repeat\_region 6813..6940  
/rpt\_family="L1"  
repeat\_region 7766..7931  
/rpt\_family="Alu"  
repeat\_region 7936..7993  
/rpt\_family="Alu"  
repeat\_region 8074..8147  
/rpt\_family="MIR"  
repeat\_region 8414..8600  
/rpt\_family="MER1\_type"  
repeat\_region 10179..10467  
/rpt\_family="Alu"  
repeat\_region 10701..10748  
/rpt\_family="(CA)n"  
repeat\_region 12532..12832  
/rpt\_family="Alu"  
repeat\_region 15471..15767  
/rpt\_family="Alu"  
repeat\_region 17166..17332  
/rpt\_family="L1"  
repeat\_region 17333..17682  
/rpt\_family="MER1\_type"  
repeat\_region 17683..17901  
/rpt\_family="L1"  
repeat\_region 18800..18863  
/rpt\_family="(GA)n"  
repeat\_region 19392..19510  
/rpt\_family="(TA)n"  
repeat\_region 19794..19912  
/rpt\_family="(TA)n"  
repeat\_region 19936..20263  
/rpt\_family="Alu"  
repeat\_region 20545..20848  
/rpt\_family="Alu"  
repeat\_region 21727..21800  
/rpt\_family="L2"  
repeat\_region 21975..22247  
/rpt\_family="Alu"  
22694..22971  
/evidence=not\_experimental  
/db\_xref="gi:1923052"  
24176..24665  
/rpt\_family="MER21\_g"  
24666..24963  
/rpt\_family="Alu"  
24968..25062  
/rpt\_family="MER21\_g"  
25466..25764  
/rpt\_family="Alu"  
25856..26163  
/rpt\_family="Alu"  
27018..27315  
/rpt\_family="Alu"  
complement(28045..28370)  
/evidence=not\_experimental

```
gene      /db_xref="GI:1235450"
complement(<28238..>117399)
/gene="WUGSC:H_DJ0587D13.1"
CDS       complement(join(<28238..28376,74985..75128,76488..76707,
82180..82351,84312..84497,109316..109564,117207..>117399))
/gene="WUGSC:H_DJ0587D13.1"
/note="atrophin-1 interacting protein 1; match to AF038563
(PID:g2947232); H_DJ0587D13.1"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC2438.1"
/db_xref="GI:3192567"
/translation="TLIGAVTAMGSADRGRLHPGDELVYVDGIPVAGKTRHYVTDLM
HNAARNGVNTVTRKVLCCGPEPCNGSPGSGVTHSSPSRSDYATVTTNNAAPSS
HNAAPGFAVSHLSQTSVVIRKENEGEGFVLIISLNRPSGSTIIVPHKIGRIIDGS
PADRCALKLVGRILLAVNGOSTINNPHADIVKLIDAGLSVTLRIPOEELNSPTSPAR
SSEKSPHAKQSLAQSPSPATNSPIAQAPQAPQLOLOQHENSYSSEVKARQ
DVKPDIPFPFDYRQPPDYRQPPGDTQPPDYRQPPPLDTRQHSPTDTRQPLS
DYRQPFDFYVDHKEGAKGFGSIRGGREYKMDLYVLRUAEDGPAIRNGMRVGDQ
IIEINGSDTRDMTHARATIELIKSGRRVRLLLKRGTGQVPEY"
repeat_region 29919..30152
/rpt_family="L2"
repeat_region 30568..30628
/rpt_family="L2"
repeat_region 30794..30958
/rpt_family="L2"
repeat_region 32619..32738
/rpt_family="L2"
repeat_region 33603..33725
/rpt_family="L2"
repeat_region 34538..34832
/rpt_family="Alu"
repeat_region 35533..35827
/rpt_family="Alu"
repeat_region 35924..36226
/rpt_family="Alu"
repeat_region 36269..36430

Query Match      1.9%; Score 20; DB 85; Length 123397;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 594 catcaaggaagacaagattt 613
|||||
Db 44441 CATCAAGGAAGACAAGATT 44460

RESULT 19
AB045363
LOCUS      AB045363 134222 bp DNA PRI 04-JUL-2000
DEFINITION Homo sapiens genomic DNA, chromosome 1q22-q23, clone:527123,
complete sequence.
ACCESSION  AB045363
VERSION    AB045363.1 GI:8918547
KEYWORDS   HTG.
SOURCE     Homo sapiens DNA, clone_11b:PAC RPC14 and 5 clone:527123.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Shilina,T., Ando,A., Takishima,N., Kikkawa,E., Iwata,K.,
Kitamura,Y., Kuwano,Y., Yamazaki,M., Soeda,E. and Inoko,H.
A 1,139,684 bp region encompassing CD1 genes on human chromosome
1q22-23
Unpublished (2000)
REFERENCE   2 (bases 1 to 134222)
AUTHORS    Shilina,T.
TITLE      Direct Submission
JOURNAL    Submitted (30-JUN-2000) to the DDBJ/EMBL/GenBank databases. Takashi
Shilina, Tokai University School of Medicine, Molecular Life Science
2; Honseidai, Isehara, Kanagawa 259-1193, Japan
(E-mail:tshilina@is.icc.u-tokai.ac.jp, Tel:81-463-93-1121(ex.2582),
Fax:81-463-94-8884)
```

```
FEATURES             Location/Qualifiers
     source            1..134222
     organism="Homo sapiens"
     db_xref="taxon:9606"
     chromosome="1"
     clone="527123"
     clone_lib="PAC RPC14 and 5"
     map="1q22-q23"
BASE COUNT  41438 a 25754 c 25673 g 41357 t
ORIGIN

Query Match      1.9%; Score 20; DB 85; Length 134222;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 778 aaatgattaaatgccactat 797
|||||
Db 13165 AAATGATTAAATGCCACTAT 13184

RESULT 20
HS3D11      146358 bp DNA PRI 23-NOV-1999
LOCUS      Human DNA sequence from clone 3D11 on chromosome Xq22.1-23 Contains
DEFINITION pseudogene similar to cytokeratin 18, EST, STS, GSS, complete
sequence.
ACCESSION  AL035088
VERSION    AL035088.1 GI:4140339
KEYWORDS   HTG; pseudogene.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 146358)
Howden,P.
Direct Submission
Submitted (07-JAN-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
IMPORTANT: This sequence is not the entire insert of clone 3D11. It
may be shorter because we only sequence overlapping sections once,
or longer because we arrange for a small overlap between
neighbouring submissions.
The true right end of clone DJ320J15 is at 100 in this sequence.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre Chromosome X
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/chrX
3D11 is from the library RPC11 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/ VECTOR: pcYPAC2.

FEATURES             Location/Qualifiers
     source            1..146358
     organism="Homo sapiens"
     db_xref="taxon:9606"
     chromosome="X"
     map="q22.1-23"
     clone="RP1-3D11"
     clone_lib="RPC1-1"
repeat_region 2..135
```

```

/notes="HAL1 repeat: matches 156. .291 of consensus"
141. .392
/notes="LIMA4 repeat: matches 6001. .6260 of consensus"
401. .574
/notes="MT1A2 repeat: matches 174. .346 of consensus"
603. .642
/notes="20 copies 2 mer gt 100% conserved"
659. .832
/notes="MT1A2 repeat: matches 3. .175 of consensus"
907. .1024
/notes="FLAM_C repeat: matches 7. .124 of consensus"
1810. .2135
/notes="MT1H repeat: matches 194. .505 of consensus"
2236. .2295
/notes="30 copies 2 mer tg 85% conserved"
2297. .2372
/notes="HAL1 repeat: matches 71. .146 of consensus"
2322. .2885
/notes="match: STS AL022461 clone 186H21"
3686. .3997
/notes="AluJb repeat: matches 1. .312 of consensus"
4007. .4536
/notes="LIPB3 repeat: matches 5630. .6150 of consensus"
4537. .4847
/notes="AluSc repeat: matches 1. .309 of consensus"
4848. .6002
/notes="LIPB3 repeat: matches 4435. .5630 of consensus"
6003. .6604
/notes="MT2D repeat: matches 1. .553 of consensus"
6605. .8036
/notes="LIPB3 repeat: matches 3012. .4439 of consensus"
8029. .8907
/notes="LIM2 repeat: matches 380. .1267 of consensus"
9067. .9207
/notes="LIM2 repeat: matches -62. .90 of consensus"
9879. .10164
/notes="AluSq repeat: matches 1. .299 of consensus"
10519. .11012
/notes="L2 repeat: matches 123. .670 of consensus"
11215. .11506
/notes="AluSc repeat: matches 1. .296 of consensus"
11556. .12044
/notes="match: GSS AQ225531"
11574. .11937
/notes="match: GSS B52954 clone 2007N8"
11600. .11748
/notes="MT1J repeat: matches 1. .145 of consensus"
complement(11732. .12059)
11896. .12016
/notes="FLAM_A repeat: matches 1. .121 of consensus"
12165. .12312
/notes="MT1D repeat: matches 1. .131 of consensus"
12822. .13176
/notes="MT1D repeat: matches 126. .505 of consensus"
13180. .13419
/notes="MIR repeat: matches 12. .255 of consensus"
13599. .13684
/notes="L2 repeat: matches 2668. .2750 of consensus"
14460. .14737
/notes="LIME repeat: matches 5529. .5819 of consensus"
14845. .14943
/notes="L2 repeat: matches 2640. .2750 of consensus"
15110. .15242
/notes="FLAM_A repeat: matches 1. .133 of consensus"
15348. .15453
/notes="MIR repeat: matches 93. .196 of consensus"
15799. .15866
/notes="34 copies 2 mer tc 68% conserved"
16873. .17045
/notes="MIR repeat: matches 93. .262 of consensus"
17989. .18279
/notes="AluX repeat: matches 3. .293 of consensus"

repeat_region
/notes="MT1E repeat: matches 309. .568 of consensus"
18354. .18602
/notes="LIPB2 repeat: matches 1146. .6146 of consensus"
23895. .23990
/notes="MT1E repeat: matches 1. .81 of consensus"
23991. .24213
/notes="HAL1 repeat: matches 276. .527 of consensus"
24257. .24404
/notes="LIME1 repeat: matches 4441. .4592 of consensus"
24446. .25410
/genes="dJ3D11.1"
/genes="dJ3D11.1"
/notes="match: x12881; M26326; match: multiple ESTs; match:
R98428 AA156469 AA132109 AI080485 AA946716 AA613797
AA811779 AA404622 AA877922 AA837881 AA160618 AA534019
AA133627 AA313627 AA204792 AA115674 AA313656 AA079316
AA428822 W45228 AA311105 AA989458 AA307774 AA642196
AA508843 AA428664 AA169563 R98427 AA044589 AA315060
AA115613 AA631699 AA404740 AA314409 AA580224 AA627959
AA334292 AA132593 AA613571 AA702833 AA007468 AA420595
AA132847 AA885759 AA100702 AA079272 AA075474 AA826722
AA287253 T87390 AA115796 AA578009 AA132800 AA192298
AA334114 AA099924 AA664179 AA826818 AA601065 AI081837
AA553886 T69875 AA577558 AA171652 AA314439 AA102231
AA857786 AA738107 AA420596 AA640574 AA224983 AA088444
AA610501 AA130711 AA115797 AA602877 AA099374 AA826349
D79150 AI198521 AA946942 AA586737 AA420528 AA134436 W45342
D794930 AA313655 AA115673 AA126394 AA160111 AA131023
AA160929 AA146738 AA829451 AA506774"
/pseudo
/codon_start=1
/evidence=not_experimental
/product="dJ3D11.1 (similar to cytokeratin 18)"
24447. .24516
/notes="LIME1 repeat: matches 4597. .4667 of consensus"
complement(24941. .>25283)
/notes="match: GSS AQ000082 clone 2282N19"
<25039. .>25434
/notes="match: GSS AQ082461 clone R-56E11"
25414. .25447
/notes="17 copies 2 mer aa 82% conserved"
25449. .25643
/notes="LIME1 repeat: matches 4655. .4834 of consensus"
25644. .25947
/notes="AluX repeat: matches 1. .303 of consensus"
25948. .26380
/notes="LIME1 repeat: matches 4834. .5233 of consensus"
26381. .26489
/notes="FLAM_C repeat: matches 1. .125 of consensus"
26490. .27290
/notes="LIME1 repeat: matches 5233. .6163 of consensus"
27309. .27383
/notes="LIME1 repeat: matches 5862. .5938 of consensus"
27386. .27572
/notes="ORSL repeat: matches 206. .400 of consensus"
27971. .28228
/notes="HAL1 repeat: matches 1444. .1709 of consensus"
28232. .28421
/notes="L2 repeat: matches 2514. .2701 of consensus"
29269. .29467
/notes="LIPAS repeat: matches 5947. .6145 of consensus"
30440. .30673
/notes="MIR repeat: matches 2. .262 of consensus"
30823. .31115
/notes="AluSq repeat: matches 1. .305 of consensus"
31181. .31513
/notes="LIMA4 repeat: matches 5968. .6300 of consensus"
31571. .31649
/notes="MIR repeat: matches 77. .152 of consensus"
31943. .32114
/notes="86 copies 2 mer ta 67% conserved"

```

```
repeat_region 33123..33453
/Note="L2 repeat: matches 1367. .1724 of consensus"
repeat_region 33457..33670
/Note="MIR repeat: matches 30. .261 of consensus"

Query Match 1.9%; Score 20; DB 92; Length 146358;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 402 ctgcccagtggtcacaaagc 421
|||||
Db 35388 CTTGCCAGTGTCAAAAGC 35407

RESULT 21
AC068306/c DNA HTG 17-AUG-2000
LOCUS Homo sapiens chromosome 3 clone RP11-202M19, WORKING DRAFT
DEFINITION AC068306
ACCESSION AC068306.12 GI:9802647
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 146463)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowle,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,E., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesha,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondajewski,N., Kong,X., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
Taber,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 146463)
Worley,K.C.
Direct Submission
Submitted (02-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Aug 15, 2000 this sequence version replaced gi:9800561.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HBBH
Center clone name: RP11-202M19
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 127593 bases at least Q40
Consensus quality: 135107 bases at least Q30
Consensus quality: 139443 bases at least Q20
Estimated insert size: 140046; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.8x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 20420: contig of 20420 bp in length
* 20421 20520: gap of unknown length
* 20521 45480: contig of 24960 bp in length
* 45481 45580: gap of unknown length
* 45581 63334: contig of 17754 bp in length
* 63335 63434: gap of unknown length
* 63435 83027: contig of 19593 bp in length
* 83028 83127: gap of unknown length
* 83128 99456: contig of 16329 bp in length
* 99457 99557: gap of unknown length
* 99558 110199: contig of 10642 bp in length
* 110200 110299: gap of unknown length
* 110300 118179: contig of 7781 bp in length
* 118180 126744: contig of 8565 bp in length
* 126745 126844: gap of unknown length
* 126845 131106: contig of 4262 bp in length
* 131107 135210: contig of 4004 bp in length
* 135211 135211: gap of unknown length
* 135212 138933: contig of 3623 bp in length
* 138934 139034: gap of unknown length
* 139035 140774: contig of 1741 bp in length
* 140775 140874: gap of unknown length
* 140875 143590: contig of 2716 bp in length
* 143591 143690: gap of unknown length
* 143691 145202: contig of 1512 bp in length
* 145203 145302: gap of unknown length
* 145303 146463: contig of 1161 bp in length.

FEATURES
Location/Qualifiers
1..146463
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-202M19"

BASE COUNT 40544 a 32550 c 31332 g 40621 t 1416 others
ORIGIN

Query Match 1.9%; Score 20; DB 73; Length 146463;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 593 tcatacgaaggaagacagatt 612
|||||
Db 84657 TCATCAGGAGACAGATT 84638

RESULT 22
AL355502
LOCUS Homo sapiens chromosome 6 clone RP11-477B18, *** SEQUENCING IN
DEFINITION AL355502
ACCESSION AL355502
VERSION AL355502.6 GI:9931755
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150152)
AUTHORS Sims,S.
```

TITLE  
JOURNAL

Direct Submission  
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
requests: clonerequest@sanger.ac.uk  
On Aug 28, 2000 this sequence version replaced gi:9863722.

## COMMENT

----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: BA477B18  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 139434 bases at least Q40  
Consensus quality: 142362 bases at least Q30  
Consensus quality: 144491 bases at least Q20  
Insert size: 147952; sum-of-contigs  
Insert size: 175899; 0.8% error; agarose-fp  
Quality coverage: 3.55x in Q20 bases; sum-of-contigs Quality  
coverage: 3.37x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 23 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 3553: contig of 3553 bp in length  
\* 3554 3653: gap of 100 bp  
\* 3654 9751: contig of 6098 bp in length  
\* 9752 9851: gap of 100 bp  
\* 9852 17450: contig of 7599 bp in length  
\* 17451 17550: gap of 100 bp  
\* 17551 22901: contig of 5351 bp in length  
\* 22902 23001: gap of 100 bp  
\* 23002 27484: contig of 4483 bp in length  
\* 27485 27584: gap of 100 bp  
\* 27585 30159: contig of 2575 bp in length  
\* 30160 30259: gap of 100 bp  
\* 30260 33956: contig of 3697 bp in length  
\* 33957 34056: gap of 100 bp  
\* 34057 37859: contig of 3803 bp in length  
\* 37860 37959: gap of 100 bp  
\* 37960 43058: contig of 5099 bp in length  
\* 43059 43158: gap of 100 bp  
\* 43159 50268: contig of 7110 bp in length  
\* 50269 50368: gap of 100 bp  
\* 50369 56506: contig of 6138 bp in length  
\* 56507 56606: gap of 100 bp  
\* 56607 69070: contig of 12464 bp in length  
\* 69071 69170: gap of 100 bp  
\* 69171 81994: contig of 12824 bp in length  
\* 81995 82094: gap of 100 bp  
\* 82095 85189: contig of 3095 bp in length  
\* 85190 85289: gap of 100 bp  
\* 85290 94643: contig of 9354 bp in length  
\* 94644 94743: gap of 100 bp  
\* 94744 97325: contig of 2582 bp in length  
\* 97326 97425: gap of 100 bp  
\* 97426 106940: contig of 9515 bp in length  
\* 106941 107040: gap of 100 bp  
\* 107041 116218: contig of 9178 bp in length  
\* 116219 116318: gap of 100 bp  
\* 116319 119546: contig of 3228 bp in length  
\* 119547 119646: gap of 100 bp  
\* 119647 122474: contig of 2828 bp in length  
\* 122475 122574: gap of 100 bp  
\* 122575 129655: contig of 7081 bp in length

## FEATURES

source  
1..150152  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP11-477B18"  
/clone\_lib="RPC1-11.2"  
1..3553  
/note="assembly\_fragment:01614  
fragment\_chain:1  
clone\_end:T7  
vector\_side:left"  
3654..9751  
/note="assembly\_fragment:01503  
fragment\_chain:1"  
9852..17450  
/note="assembly\_fragment:01244  
fragment\_chain:1"  
17551..22901  
/note="assembly\_fragment:01044  
fragment\_chain:1"  
23002..27484  
/note="assembly\_fragment:01629  
fragment\_chain:2"  
27585..30159  
/note="assembly\_fragment:01631  
fragment\_chain:2"  
30260..33956  
/note="assembly\_fragment:00502  
fragment\_chain:2"  
34057..37859  
/note="assembly\_fragment:01432  
fragment\_chain:2"  
37960..43058  
/note="assembly\_fragment:00405  
fragment\_chain:2"  
43159..50268  
/note="assembly\_fragment:01016  
fragment\_chain:3"  
50369..56506  
/note="assembly\_fragment:00860  
fragment\_chain:3"  
56607..69070  
/note="assembly\_fragment:00437  
fragment\_chain:3"  
69171..81994  
/note="assembly\_fragment:01040  
fragment\_chain:3"  
82095..85189  
/note="assembly\_fragment:00091  
fragment\_chain:4"  
85290..94643  
/note="assembly\_fragment:00369  
fragment\_chain:4"  
94744..97325  
/note="assembly\_fragment:00207  
fragment\_chain:4"  
97426..106940  
/note="assembly\_fragment:00844  
fragment\_chain:5"  
107041..116218  
/note="assembly\_fragment:00840  
fragment\_chain:5"  
116319..119546  
/note="assembly\_fragment:00452  
fragment\_chain:5"  
119647..122474  
/note="assembly\_fragment:01555  
fragment\_chain:6"

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature 122575. .129655  
/note="assembly\_fragment:00103  
fragment\_chain:6"  
misc\_feature 129756. .139583  
/note="assembly\_fragment:01041"  
misc\_feature 138684. .150152  
/note="assembly\_fragment:01043"  
BASE COUNT 42791 a 29791 c 30984 g 44377 t 2209 others  
ORIGIN

Query Match 1.9%; Score 20; DB 79; Length 150152;  
Best Local Similarity 100.0%; Pred. No. 9.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 agtactactgtgtccaaatc 306  
|||||  
Db 98661 AGTACTACGTGTCCAAATC 98680

RESULT 23  
AC012545/c  
LOCUS  
DEFINITION Homo sapiens clone RP11-3B6, WORKING DRAFT SEQUENCE, 4 unordered  
pieces.  
VERSION AC012545  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 152719)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Unpublished  
2 (bases 1 to 152719)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,  
Brown,A., Castie,A., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,  
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
McEwan,P., McGurk,A., McKernan,K., O'Connor,T., O'Donnell,P.,  
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (29-OCT-1999), Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Nov 3, 2000 this sequence version replaced gi:10045323.  
All repeats were identified using RepeatMasker:  
Snit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L2706  
Center clone name: 3\_B6  
----- Summary Statistics  
Sequencing vector: M13: M77815; 50% of reads  
Sequencing vector: Plasmid: n/a; 50% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 152006 bases at least Q40  
Consensus quality: 152242 bases at least Q30

Consensus quality: 152343 bases at least Q20  
Insert size: 148000; agarose-fp  
Insert size: 152419; sum-of-contigs  
Quality coverage: 11.6 in Q20 bases; agarose-fp  
Quality coverage: 11.2 in Q20 ba.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 50537: contig of 50537 bp in length  
\* 50538 50637: gap of 100 bp  
\* 50638 51266: contig of 629 bp in length  
\* 51267 51366: gap of 100 bp  
\* 51367 52001: contig of 635 bp in length  
\* 52002 52101: gap of 100 bp  
\* 52102 152719: contig of 100618 bp in length.  
FEATURES  
Location/Qualifiers  
1. .152719  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-3B6"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .50537  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
50638..51266  
/note="assembly\_fragment"  
51367..52001  
/note="assembly\_fragment"  
52102..152719  
/note="assembly\_fragment  
clone\_end:T7  
vector\_side:right"  
BASE COUNT 46108 a 27402 c 29352 g 49557 t 300 others  
ORIGIN

Query Match 1.9%; Score 20; DB 62; Length 152719;  
Best Local Similarity 100.0%; Pred. No. 9.3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 tactcaaaacatgaagata 643  
|||||  
Db 113683 TACTCAAAACATGAGGATA 113664

RESULT 24  
AL513208  
LOCUS  
DEFINITION Homo sapiens chromosome 1 clone RP11-393K10, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 15 unordered pieces.  
AL513208  
ACCESSION AL513208.3 GI:12597159  
VERSION  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 153487)  
Pavitt,R.  
Direct Submission  
Submitted (27-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerquest@sanger.ac.uk  
On Jan 29, 2001 this sequence version replaced gi:12584793.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC

Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 ----- Project Information  
 Center project name: BA393K10  
 ----- Summary Statistics

Assembly program: XGAP4; version 4.5  
 Sequencing vector: piasmld; L08752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 147995 bases at least Q40  
 Consensus quality: 150052 bases at least Q30  
 Consensus quality: 151312 bases at least Q20  
 Insert size: 152087; sum-of-contigs  
 Insert size: 157476; 4.9% error; agarose-fp  
 Quality coverage: 4.29x in Q20 bases; sum-of-contigs Quality  
 coverage: 4.28x in Q20 bases; agarose-fp  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 15 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 5666: contig of 5666 bp in length  
 \* 5667 5766: gap of 100 bp  
 \* 5767 26593: contig of 20827 bp in length  
 \* 26594 26693: gap of 100 bp  
 \* 26694 28820: contig of 2127 bp in length  
 \* 28821 28920: gap of 100 bp  
 \* 28921 31730: contig of 2810 bp in length  
 \* 31731 31830: gap of 100 bp  
 \* 31831 34525: contig of 2695 bp in length  
 \* 34526 34625: gap of 100 bp  
 \* 34626 42590: contig of 7965 bp in length  
 \* 42591 42690: gap of 100 bp  
 \* 42691 48846: contig of 6156 bp in length  
 \* 48847 48946: gap of 100 bp  
 \* 48947 76902: contig of 27956 bp in length  
 \* 76903 77002: gap of 100 bp  
 \* 77003 82856: contig of 5854 bp in length  
 \* 82857 82956: gap of 100 bp  
 \* 82957 98969: contig of 16013 bp in length  
 \* 98970 99069: gap of 100 bp  
 \* 99070 120785: contig of 21716 bp in length  
 \* 120786 120885: gap of 100 bp  
 \* 120886 134804: contig of 13919 bp in length  
 \* 134805 134904: gap of 100 bp  
 \* 134905 139191: contig of 4287 bp in length  
 \* 139192 139291: gap of 100 bp  
 \* 139292 142279: contig of 2988 bp in length  
 \* 142280 142379: gap of 100 bp  
 \* 142380 153487: contig of 11108 bp in length.

#### FEATURES

source  
 1. .153487  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosomes="1"  
 /clone="RP11-393K10"  
 /clone\_lib="RPC1-11.2"  
 1. .5666  
 /note="assembly\_fragment:01066  
 fragment\_chain:1"  
 5767..26593  
 /note="assembly\_fragment:00944  
 fragment\_chain:1"  
 26694..28820  
 /note="assembly\_fragment:01168  
 fragment\_chain:1"  
 28921..31730  
 /note="assembly\_fragment:00745  
 fragment\_chain:1"  
 31831..34525

misc\_feature  
 /note="assembly\_fragment:00937  
 fragment\_chain:2"  
 34626..42590  
 /note="assembly\_fragment:00870  
 fragment\_chain:2"  
 42691..48846  
 /note="assembly\_fragment:01337  
 fragment\_chain:3"  
 48947..76902  
 /note="assembly\_fragment:00956  
 fragment\_chain:3"  
 77003..82856  
 /note="assembly\_fragment:00750"  
 82957..98969  
 /note="assembly\_fragment:00368  
 fragment\_chain:4"  
 99070..120785  
 /note="assembly\_fragment:00251  
 fragment\_chain:4"  
 120886..134804  
 /note="assembly\_fragment:00833  
 fragment\_chain:4"  
 134905..139191  
 /note="assembly\_fragment:00894  
 fragment\_chain:4"  
 139292..142279  
 /note="assembly\_fragment:00107  
 fragment\_chain:4"  
 142380..153487  
 /note="assembly\_fragment:01096  
 fragment\_chain:4  
 clone\_end:SP6  
 vector\_side:right"  
 BASE COUNT 45722 a 28611 c 28538 g 49214 t 1402 others  
 ORIGIN  
 Query Match 1.9%; Score 20; DB 81; Length 153487;  
 Best Local Similarity 100.0%; Pred. No. 9.3;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 778 aaatggattaatgccactat 797  
 |||||  
 Db 46470 AAATGGATTAAATGCCACTAT 46489  
 RESULT 25  
 AC006466 156880 bp DNA PRI 02-OCT-2000  
 LOCUS Homo sapiens BAC clone CTA-471E18 from 7p11.2-p21, complete  
 DEFINITION sequence.  
 AC006466  
 AC006466.3 GI:5836158  
 VERSION HTG.  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 156880)  
 Sulston, J.E. and Waterston, R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 2 (bases 1 to 156880)  
 Edwards, J., Wohlmann, P. and Duckels, G.  
 TITLE The sequence of Homo sapiens BAC clone CTA-471E18  
 JOURNAL Unpublished  
 3 (bases 1 to 156880)  
 Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-JAN-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA





```

repeat_region 24075..24168
misc_feature   /rpt_family="MIR"
misc_feature   24354..24468
misc_feature   /note="match to EST AI815550 (NID:g5431096) au48b07.y1"
misc_feature   24397..24451
misc_feature   /note="match to EST T08815 (NID:g389843)"
repeat_region 25002..25138
repeat_region /rpt_family="MERL_type"
repeat_region 26122..26286
repeat_region /rpt_family="MERL_type"
repeat_region 26965..27145
repeat_region /rpt_family="MIR"
repeat_region 27224..27399
repeat_region /rpt_family="MIR"
repeat_region 27719..27765
repeat_region /rpt_family="MIR"
repeat_region 27769..27957
repeat_region /rpt_family="MIR"
misc_feature 27915..28072
misc_feature /note="match to EST AA077136 (NID:g1836610)"
repeat_region 28506..28665
repeat_region /rpt_family="MERL_type"
repeat_region 30255..30420
repeat_region /rpt_family="MERL_type"
repeat_region 31077..31478
repeat_region /rpt_family="L2"
repeat_region 31479..31553
repeat_region /rpt_family="(TCCA)n"
repeat_region 31554..31620

Query Match 1.9%; Score 20; DB 86; Length 156880;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 939 acaggctgggaagaagaatg 958
|||||
Db 155732 ACAGGCTGGGAAGAAGATG 155751

RESULT 26
AC055782/c
LOCUS AC055782 159235 bp DNA HTG 13-MAY-2000
DEFINITION Homo sapiens chromosome 15 clone RP11-528P19 map 15, WORKING DRAFT
SEQUENCE, 30 unordered pieces.
AC055782
VERSION AC055782.2 GI:7798784
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159235)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
Campyano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Gadigan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

```

# TITLE JOURNAL COMMENT

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On May 13, 2000 this sequence version replaced gi:7582606.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L6233  
 Center clone name: 528\_P\_19  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 143511 bases at least Q40  
 Consensus quality: 151834 bases at least Q30  
 Consensus quality: 154789 bases at least Q20  
 Insert size: 182000; agarose-fp  
 Insert size: 156335; sum-of-contigs  
 Quality coverage: 3.3 in Q20 bases; agarose-fp  
 Quality coverage: 3.8 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 30 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1233: contig of 1233 bp in length  
 \* 1234 1333: gap of 100 bp  
 \* 1334 1341: contig of 208 bp in length  
 \* 1542 1641: gap of 100 bp  
 \* 1642 2954: contig of 1313 bp in length  
 \* 2955 3054: gap of 100 bp  
 \* 3055 4487: contig of 1433 bp in length  
 \* 4488 4587: gap of 100 bp  
 \* 4588 7019: contig of 2432 bp in length  
 \* 7020 7119: gap of 100 bp  
 \* 7120 9033: contig of 1914 bp in length  
 \* 9034 9133: gap of 100 bp  
 \* 9134 11313: contig of 2180 bp in length  
 \* 11314 11413: gap of 100 bp  
 \* 11414 14949: contig of 3536 bp in length  
 \* 14950 15049: gap of 100 bp  
 \* 15050 18020: contig of 2971 bp in length  
 \* 18021 18120: gap of 100 bp  
 \* 18121 21534: contig of 3414 bp in length  
 \* 21535 21634: gap of 100 bp  
 \* 21635 24294: contig of 2660 bp in length  
 \* 24295 24394: gap of 100 bp  
 \* 24395 27309: contig of 2915 bp in length  
 \* 27310 27409: gap of 100 bp  
 \* 27410 29656: contig of 2247 bp in length  
 \* 29657 29756: gap of 100 bp  
 \* 29757 33306: contig of 3550 bp in length  
 \* 33307 33406: gap of 100 bp  
 \* 33407 35718: contig of 2312 bp in length  
 \* 35719 35818: gap of 100 bp  
 \* 35819 40112: contig of 4294 bp in length  
 \* 40113 40212: gap of 100 bp  
 \* 40213 44197: contig of 3985 bp in length  
 \* 44198 44297: gap of 100 bp

```
* 44298 48353: contig of 4056 bp in length
* 48354 48453: gap of 100 bp
* 48454 55082: contig of 6629 bp in length
* 55083 55182: gap of 100 bp
* 55183 60528: contig of 5346 bp in length
* 60529 60628: gap of 100 bp
* 60629 66609: contig of 5981 bp in length
* 66610 66709: gap of 100 bp
* 66710 73843: contig of 7134 bp in length
* 73844 73943: gap of 100 bp
* 73944 81343: contig of 7400 bp in length
* 81344 81443: gap of 100 bp
* 81444 90270: contig of 8827 bp in length
* 90271 90370: gap of 100 bp
* 90371 98386: contig of 8016 bp in length
* 98387 98486: gap of 100 bp
* 98487 105549: contig of 7063 bp in length
* 105550 105649: gap of 100 bp
* 105650 116400: contig of 10751 bp in length
* 116401 116500: gap of 100 bp
* 116501 129205: contig of 12705 bp in length
* 129206 129305: gap of 100 bp
* 129306 144057: contig of 14752 bp in length
* 144058 144157: gap of 100 bp
* 144158 159235: contig of 15078 bp in length.
```

FEATURES

```
source
1..159235
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="15"
/map="15"
/clone="RP11-528p19"
/clone_lib="RPCI-11 Human Male BAC"
1..1233
/note="assembly_fragment"
1334..1541
/note="assembly_fragment"
clone_end:T7
vector_side:right
1642..2954
/note="assembly_fragment"
3055..4487
/note="assembly_fragment"
4588..7019
/note="assembly_fragment"
7120..9033
/note="assembly_fragment"
9134..11313
/note="assembly_fragment"
11414..14949
/note="assembly_fragment"
15050..18020
/note="assembly_fragment"
18121..21534
/note="assembly_fragment"
21635..24294
/note="assembly_fragment"
24395..27309
/note="assembly_fragment"
27410..29656
/note="assembly_fragment"
clone_end:SP6
vector_side:left
29757..33306
/note="assembly_fragment"
33407..35718
/note="assembly_fragment"
35819..40112
/note="assembly_fragment"
40213..44197
/note="assembly_fragment"
44298..48353
/note="assembly_fragment"
```

```
misc_feature 48454..55082
/note="assembly_fragment"
misc_feature 55183..60528
/note="assembly_fragment"
misc_feature 60629..66609
/note="assembly_fragment"
misc_feature 66710..73843
/note="assembly_fragment"
misc_feature 73944..81343
/note="assembly_fragment"
misc_feature 81444..90270
/note="assembly_fragment"
misc_feature 90371..98386
/note="assembly_fragment"
misc_feature 98487..105549
/note="assembly_fragment"
misc_feature 105650..116400
/note="assembly_fragment"
misc_feature 116501..129205
/note="assembly_fragment"
misc_feature 129306..144057
/note="assembly_fragment"

Query Match 1.9%; Score 20; DB 72; Length 159235;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 aatgattaatgcactatt 798
|||||
Db 85080 AATGGATTAAATGCCACTATT 85061

RESULT 27
AL356482/c
LOCUS AL356482 169542 bp DNA HTG 19-JAN-2001
DEFINITION Homo sapiens chromosome 15 clone RP11-358L18, *** SEQUENCING IN
PROGRESS ***, 14 unordered pieces.
ACCESSION AL356482
VERSION AL356482.6 GI:9863778
KEYWORDS HTG; TGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 169542)
AUTHORS Sims,S.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:9213910.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA358L18
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 162328 bases at least Q40
Consensus quality: 165328 bases at least Q30
Consensus quality: 168669 bases at least Q20
Insert size: 168242; sum-of-contigs
Insert size: 181476; agarose-fp
Quality coverage: 4.03x in Q20 bases; sum-of-contigs Quality
coverage: 3.89x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
```

\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 15516: contig of 15516 bp in length  
\* 15517 15616: gap of 100 bp  
\* 15617 24420: contig of 8804 bp in length  
\* 24421 24520: gap of 100 bp  
\* 24521 47898: contig of 23378 bp in length  
\* 47899 47998: gap of 100 bp  
\* 47999 51313: contig of 3315 bp in length  
\* 51314 51413: gap of 100 bp  
\* 51414 55861: contig of 4448 bp in length  
\* 55862 55961: gap of 100 bp  
\* 55962 60639: contig of 4678 bp in length  
\* 60640 60739: gap of 100 bp  
\* 60740 66399: contig of 5660 bp in length  
\* 66400 66499: gap of 100 bp  
\* 66500 79258: contig of 12759 bp in length  
\* 79259 79358: gap of 100 bp  
\* 79359 84616: contig of 5258 bp in length  
\* 84617 84716: gap of 100 bp  
\* 84717 88488: contig of 3772 bp in length  
\* 88489 88588: gap of 100 bp  
\* 88589 128005: contig of 39417 bp in length  
\* 128006 128105: gap of 100 bp  
\* 128106 139254: contig of 11149 bp in length  
\* 139255 139354: gap of 100 bp  
\* 139355 165798: contig of 26444 bp in length  
\* 165799 165898: gap of 100 bp  
\* 165899 169542: contig of 3644 bp in length.

## FEATURES

## source

1. .169542  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/clone="RP11-358L18"  
/clone\_lib="RPC1-11.2"  
1. .15516  
/note="assembly\_fragment:00900  
fragment\_chain:1"  
15617. .24420  
/note="assembly\_fragment:00976  
fragment\_chain:1"  
24521. .47898  
/note="assembly\_fragment:01099  
fragment\_chain:2"  
47999. .51313  
/note="assembly\_fragment:01341  
fragment\_chain:2"  
51414. .55861  
/note="assembly\_fragment:01450  
fragment\_chain:3"  
55962. .60639  
/note="assembly\_fragment:00032  
fragment\_chain:3"  
60740. .66399  
/note="assembly\_fragment:00117"  
66500. .79258  
/note="assembly\_fragment:00167"  
79359. .84616  
/note="assembly\_fragment:00225"  
84717. .88488  
/note="assembly\_fragment:01331"  
88589. .128005  
/note="assembly\_fragment:01386"  
128106. .139254  
/note="assembly\_fragment:01627"  
139355. .165798  
/note="assembly\_fragment:00559  
fragment\_chain:4"  
165899. .169542

/note="assembly\_fragment:01272  
fragment\_chain:4  
clone\_end:SP6  
vector\_side:right"

BASE COUNT 48651 a 35926 c 34828 g 48834 t 1303 others  
ORIGIN

Query Match 1.9%; Score 20; DB 80; Length 169542;  
Best Local Similarity 100.0%; Pred. No. 9, 3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 625 actcaaacatgaagatag 644  
|||||  
Db 5004 ACTCAAAACATGAAGATAG 4985

## RESULT 28

AC073132/c

LOCUS

DEFINITION

SEQUENCE, 2 unordered pieces.

AC073132

VERSION

AC073132.6 GI:11345026

KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 172148)

Waterston,R.H.

The sequence of Homo sapiens clone

Unpublished

2 (bases 1 to 172148)

Waterston,R.H.

Direct Submission

Submitted (08-JUN-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

On Nov 26, 2000 this sequence version replaced gi:11192275.

COMMENT

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc/index.shtml

----- Project Information -----

Center project name: HNH0684G24

----- Summary Statistics -----

Sequencing vector: M13; 65%

Chemistry: Dye-terminator Big Dye; 35%

Chemistry: Dye-terminator Big Dye; 35% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 171642 bases at least Q40

Consensus quality: 171824 bases at least Q30

Consensus quality: 171929 bases at least Q20

Insert size: 171000; agarose-fp

Insert size: 172048; sum-of-contigs

Quality coverage: 8.38 in Q20 bases; agarose-fp

Quality coverage: 8.77 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 2 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 42185: contig of 42185 bp in length

\* 42186 42285: gap of unknown length

\* 42286 172148: contig of 129863 bp in length.

Location/Qualifiers

FEATURES

source 1. .172148

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-684G24"  
1. .42185  
/note="assembly\_name:Contig1  
clone\_end:r7  
vector\_side:right"  
42286. .172148  
/note="assembly\_name:Contig2  
clone\_end:SP6  
vector\_side:left"

BASE COUNT 55690 a 33742 c 32633 g 49983 t 100 others  
ORIGIN

Query Match 1.9%; Score 20; DB 74; Length 172148;

Best Local Similarity 100.0%; Pred. No. 9,3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 939 acagggctgggaagagaatg 958

|||||

Db 157389 ACAGGGCTGGGAAGAGAATG 157370

RESULT 29

AL359753 180742 bp DNA PRI 01-MAR-2001  
LOCUS Human DNA sequence from clone RP11-520H16 on chromosome 1, complete  
DEFINITION

ACCESSION AL359753

VERSION AL359753.9 GI:12831858

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 180742)

Bates,K.

Direct Submission

Submitted (01-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Feb 15, 2001 this sequence version replaced GI:12718072.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em:, EMBL; Sw:,

SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP

database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping

Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr1

RP11-520H16 is from the library RPCI-11.2 constructed by the group

of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACE3.6

This sequence is the entire insert of clone RP11-520H16 The true

.left end of clone RP11-59J4 is at 138164 in this sequence. The true

right end of clone RP11-393K10 is at 101982 in this sequence.

FEATURES

Source

Location/Qualifiers

1. .180742

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/clone="RP11-520H16"

/clone\_lib="RPCI-11.2"

8. .490

/note="match: GSS: Em:AQ833358"

565. .607

/note="L1M3c repeat: matches 1094. .1135 of consensus"

608. .901

/note="AluSp repeat: matches 1. .297 of consensus"

902. .1460

/note="L1M3c repeat: matches 1135. .1297 of consensus"

1464. .1758

/note="L1PA12 repeat: matches 5661. .5960 of consensus"

4295. .4480

/note="AluSq repeat: matches 118. .308 of consensus"

4491. .6358

/note="L1 repeat: matches 3104. .4988 of consensus"

6367. .6461

/note="L1M1 repeat: matches -1389. .-1285 of consensus"

6473. .7952

/note="L1PA13 repeat: matches -305. .1154 of consensus"

7928. .7961

/note="L1PA13 repeat: matches 1307. .1339 of consensus"

7962. .8495

/note="L1PA13 repeat: matches 5623. .6155 of consensus"

9276. .10662

/note="L1 repeat: matches 2125. .3505 of consensus"

10662. .11384

/note="L1MA8 repeat: matches 5581. .6289 of consensus"

11916. .11991

/note="L2 repeat: matches 2656. .2733 of consensus"

13046. .13196

/note="FAM repeat: matches 1. .162 of consensus"

14123. .14217

/note="L1MC/D repeat: matches 5305. .5413 of consensus"

14974. .15053

/note="MIR repeat: matches 21. .110 of consensus"

15238. .15564

/note="L2 repeat: matches 2404. .2735 of consensus"

16156. .16522

/note="THE1C repeat: matches 3. .371 of consensus"

16524. .17988

/note="THE1C-internal repeat: matches 17. .1533 of

consensus"

17999. .18360

/note="THE1C repeat: matches 15. .371 of consensus"

19224. .19561

/note="MLT1A1 repeat: matches 1. .365 of consensus"

19751. .20016

/note="match: GSS: Em:AQ675721"

complement(20115. .20434)

/note="match: GSS: Em:AQ582713"

20887. .20974

/note="MIR repeat: matches 19. .107 of consensus"

20980. .21144

/note="MLT1A2 repeat: matches 1. .178 of consensus"

21146. .21237

/note="46 copies 2 mer tc 78% conserved"

21239. .21441

/note="MLT1A2 repeat: matches 173. .374 of consensus"

complement(21446. .22253)

/note="match: STS: Em:L30626"

22002. .22043

/note="14 copies 3 mer gat 85% conserved"

22046. .22292

/note="AluSq repeat: matches 38. .284 of consensus"

22379. .22423

/note="MER63 repeat: matches 676. .720 of consensus"

```
repeat_region 22733..23160
/notes="L1MEC repeat: matches 1615..2158 of consensus"
repeat_region 23246..23873
/notes="L1R9 repeat: matches 10..625 of consensus"
repeat_region 24112..24569
/notes="L1MCB repeat: matches 834..1282 of consensus"
repeat_region 24572..24932
/notes="L1M4 repeat: matches 3242..3603 of consensus"
misc_feature 24879..25247
/notes="match: GSS: Em:AQ170065"
repeat_region 25175..25337
/notes="L1M4 repeat: matches 3693..3850 of consensus"
repeat_region 25622..26007
/notes="MER578 repeat: matches 1..400 of consensus"
repeat_region 28964..29026
/notes="L1PA16 repeat: matches 5784..5840 of consensus"
repeat_region 29030..29333
/notes="L1PA16 repeat: matches 5844..6156 of consensus"
misc_feature complement(29809..30266)
/notes="match: GSS: Em:AQ212470"
repeat_region 30008..30155
/notes="37 copies 4 mer atat 60% conserved"
repeat_region 31584..31855
/notes="AluSq repeat: matches 33..307 of consensus"
repeat_region 31938..32109
/notes="L2 repeat: matches 2356..2533 of consensus"
repeat_region 32937..33233
/notes="L1PA4 repeat: matches 5850..6146 of consensus"
repeat_region 33234..33693
/notes="L1P repeat: matches 3649..4120 of consensus"
repeat_region 34533..34835
/notes="AluY repeat: matches 1..304 of consensus"
misc_feature complement(35719..36242)
/notes="match: GSS: Em:AQ319180"
repeat_region 36382..36811
/notes="L2 repeat: matches 2257..2710 of consensus"
repeat_region 36819..37123
/notes="AluSx repeat: matches 10..312 of consensus"
misc_feature 43237..43414
/notes="match: GSS: Em:AQ106632"
repeat_region 43416..43459
/notes="11 copies 4 mer taga 100% conserved"
misc_feature 43420..43578
/notes="match: GSS: Em:AQ106632"
repeat_region 44428..44920
/notes="L1R22 repeat: matches 1..505 of consensus"
repeat_region 44923..46258
/notes="HVRK22 repeat: matches 5503..6835 of consensus"
repeat_region 46259..46691
/notes="L1PA3 repeat: matches 5714..6146 of consensus"
repeat_region 46692..48529
/notes="HVRK22 repeat: matches 3655..5503 of consensus"
repeat_region 48529..48591
/notes="HVRK22 repeat: matches 2493..2545 of consensus"
repeat_region 48574..48653
/notes="HVRK22 repeat: matches 2829..2907 of consensus"
repeat_region 48653..48796
/notes="HVRK22 repeat: matches 2327..2473 of consensus"
repeat_region 49215..49607
/notes="HVRK22 repeat: matches 339..741 of consensus"
repeat_region 49609..50386
/notes="L1PA2 repeat: matches 1..776 of consensus"
repeat_region 50382..55628
/notes="L1PA2 repeat: matches 900..6146 of consensus"
repeat_region 55629..55664
/notes="18 copies 2 mer aa 86% conserved"
repeat_region 55666..56016
/notes="HVRK22 repeat: matches 3..350 of consensus"
repeat_region 56024..56512
/notes="L1R22 repeat: matches 1..505 of consensus"
misc_feature 56276..56679
/notes="match: STS: Em:G16240"
repeat_region 58282..63993
```

```
repeat_region /notes="L1PA15 repeat: matches -1416..4402 of consensus"
64119..64240
/notes="MIR repeat: matches 100..232 of consensus"
65689..65928
/notes="L1PA15 repeat: matches 5920..6157 of consensus"
67171..67342
/notes="L1PB3 repeat: matches 5969..6138 of consensus"
67364..67417
/notes="U2 repeat: matches 1..55 of consensus"

Query Match 1.9%; Score 20; DB 90; Length 180742;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 778 aaatgattaaatgcactat 797
|||||
Db 19346 AAATGGATTAAATGCCACTAT 19365

RESULT 30
AC022254/c
LOCUS AC022254
DEFINITION Homo sapiens chromosome 15 clone RP11-282M16 map 15, WORKING DRAFT
SEQUENCE, 27 unordered pieces.
ACCESSION AC022254
VERSION AC022254.2 GI:7239619
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 181505)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 15, clone RP11-282M16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 181505)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Deaellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
TITLE
JOURNAL
COMMENT
Submitted (27-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 14, 2000 this sequence version replaced gi:6778503.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5066
Center clone name: 282_M.16
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 167001 bases at least Q40
```

Consensus quality: 173687 bases at least Q30  
Consensus quality: 176383 bases at least Q20  
Insert size: 172000; agarose-fp  
Insert size: 178905; sum-of-contigs  
Quality coverage: 4.5 in Q20 bases; agarose-fp  
Quality coverage: 4.3 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 27 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1078: contig of 1078 bp in length  
\* 1079 1178: gap of 100 bp  
\* 1179 2268: contig of 1090 bp in length  
\* 2269 2368: gap of 100 bp  
\* 2369 3728: contig of 1360 bp in length  
\* 3729 3828: gap of 100 bp  
\* 3829 5984: contig of 2156 bp in length  
\* 5985 6084: gap of 100 bp  
\* 6085 7767: contig of 1683 bp in length  
\* 7768 7867: gap of 100 bp  
\* 7868 10416: contig of 2549 bp in length  
\* 10417 10516: gap of 100 bp  
\* 10517 12313: contig of 1797 bp in length  
\* 12314 12413: gap of 100 bp  
\* 12414 15021: contig of 2608 bp in length  
\* 15022 15121: gap of 100 bp  
\* 15122 17839: contig of 2718 bp in length  
\* 17840 17939: gap of 100 bp  
\* 17940 21118: contig of 3179 bp in length  
\* 21119 21218: gap of 100 bp  
\* 21219 25195: contig of 3977 bp in length  
\* 25196 25295: gap of 100 bp  
\* 25296 30010: contig of 4715 bp in length  
\* 30011 30110: gap of 100 bp  
\* 30111 34082: contig of 3972 bp in length  
\* 34083 34182: gap of 100 bp  
\* 34183 40143: contig of 5961 bp in length  
\* 40144 40243: gap of 100 bp  
\* 40244 46062: contig of 5819 bp in length  
\* 46063 46162: gap of 100 bp  
\* 46163 51764: contig of 5602 bp in length  
\* 51765 51864: gap of 100 bp  
\* 51865 56797: contig of 4933 bp in length  
\* 56798 56897: gap of 100 bp  
\* 56898 61834: contig of 4937 bp in length  
\* 61835 61934: gap of 100 bp  
\* 61935 68365: contig of 6431 bp in length  
\* 68366 68465: gap of 100 bp  
\* 68466 75130: contig of 6665 bp in length  
\* 75131 75230: gap of 100 bp  
\* 75231 82142: contig of 6912 bp in length  
\* 82143 82242: gap of 100 bp  
\* 82243 92535: contig of 10293 bp in length  
\* 92536 92635: gap of 100 bp  
\* 92636 101441: contig of 8806 bp in length  
\* 101442 101541: gap of 100 bp  
\* 101542 109566: contig of 8025 bp in length  
\* 109567 109666: gap of 100 bp  
\* 109667 122342: contig of 12676 bp in length  
\* 122343 122442: gap of 100 bp  
\* 122443 135713: contig of 13271 bp in length  
\* 135714 135813: gap of 100 bp  
\* 135814 181505: contig of 45692 bp in length.

FEATURES  
source

1. 181505  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"

/map="15"  
/clone="RP11-282M16"  
/db\_lib="RPC1-11 Human Male BAC"  
1. 1078  
/note="assembly\_fragment"  
misc\_feature  
1179. 2268  
/note="assembly\_fragment"  
misc\_feature  
2369. 3728  
/note="assembly\_fragment"  
misc\_feature  
3829. 5984  
/note="assembly\_fragment"  
misc\_feature  
6085. 7767  
/note="assembly\_fragment"  
misc\_feature  
7868. 10416  
/note="assembly\_fragment"  
misc\_feature  
10517. 12313  
/note="assembly\_fragment"  
misc\_feature  
12414. 15021  
/note="assembly\_fragment"  
misc\_feature  
15122. 17839  
/note="assembly\_fragment"  
misc\_feature  
17940. 21118  
/note="assembly\_fragment"  
misc\_feature  
21219. 25195  
/note="assembly\_fragment"  
misc\_feature  
25296. 30010  
/note="assembly\_fragment"  
misc\_feature  
30111. 34082  
/note="assembly\_fragment"  
misc\_feature  
clone\_end:SP6  
vector\_side:left  
34183. 40143  
/note="assembly\_fragment"  
misc\_feature  
40244. 46062  
/note="assembly\_fragment"  
misc\_feature  
46163. 51764  
/note="assembly\_fragment"  
misc\_feature  
51865. 56797  
/note="assembly\_fragment"  
misc\_feature  
56898. 61834  
/note="assembly\_fragment"  
misc\_feature  
61935. 68365  
/note="assembly\_fragment"  
misc\_feature  
68466. 75130  
/note="assembly\_fragment"  
misc\_feature  
75231. 82142  
/note="assembly\_fragment"  
misc\_feature  
82243. 92535  
/note="assembly\_fragment"  
misc\_feature  
92636. 101441  
/note="assembly\_fragment"  
misc\_feature  
101542. 109566  
/note="assembly\_fragment"  
misc\_feature  
109667. 122342  
/note="assembly\_fragment"  
misc\_feature  
122443. 135713  
/note="assembly\_fragment"  
misc\_feature  
135814. 181505  
/note="assembly\_fragment"  
misc\_feature  
BASE COUNT 48952 a 40745 c 41288 g 47915 t 2605 others  
ORIGIN

Query Match 1.9%; Score 20; DB 67; Length 181505;  
Best Local Similarity 100.0%; Pred. No. 9.3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 779 aatggattaatgcactatt 798  
|||||  
Db 34467 AATGGATTAAATGCCACTATT 34448

RESULT 31







```

* 94340 94439: gap of unknown length
* 94440 104697: contig of 10258 bp in length
* 104698 104797: gap of unknown length
* 104798 115243: contig of 10446 bp in length
* 115244 115343: gap of unknown length
* 115344 125658: contig of 10315 bp in length
* 125659 125758: gap of unknown length
* 125759 138956: contig of 13198 bp in length
* 138957 139056: gap of unknown length
* 139057 154201: contig of 15145 bp in length
* 154202 154301: gap of unknown length
* 154302 168775: contig of 14374 bp in length
* 168776 168775: gap of unknown length
* 168776 188603: contig of 19828 bp in length.

FEATURES
  source
    1..188603
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="UNK"
      /clone="RP11-484E6"
    1..1600
      /note="assembly_name:Contig19"
  misc_feature
    1701..2986
      /note="assembly_name:Contig20"
  misc_feature
    3087..8207
      /note="assembly_name:Contig28"
      clone_end="7"
      vector_side="right"
  misc_feature
    8308..11075
      /note="assembly_name:Contig21"
  misc_feature
    11176..14112
      /note="assembly_name:Contig22"
  misc_feature
    14213..17455
      /note="assembly_name:Contig23"
  misc_feature
    17556..21575
      /note="assembly_name:Contig24"
  misc_feature
    21676..25659
      /note="assembly_name:Contig25"
  misc_feature
    25760..30093
      /note="assembly_name:Contig26"
  misc_feature
    30194..36128
      /note="assembly_name:Contig27"
  misc_feature
    36229..41487
      /note="assembly_name:Contig29"
  misc_feature
    41588..47658
      /note="assembly_name:Contig30"
  misc_feature
    47759..54729
      /note="assembly_name:Contig31"
  misc_feature
    54830..59691
      /note="assembly_name:Contig32"
  misc_feature
    59792..67976
      /note="assembly_name:Contig33"
  misc_feature
    68077..75027
      /note="assembly_name:Contig34"
  misc_feature
    75128..83355
      /note="assembly_name:Contig35"
  misc_feature
    83456..94339
      /note="assembly_name:Contig36"
  misc_feature
    94440..104697
      /note="assembly_name:Contig37"
  misc_feature
    104798..115243
      /note="assembly_name:Contig38"
  misc_feature
    115344..125658
      /note="assembly_name:Contig39"
  misc_feature
    125759..138956
      /note="assembly_name:Contig40"
  misc_feature
    139057..154201
      /note="assembly_name:Contig41"
  misc_feature
    154302..168675
      /note="assembly_name:Contig42"
  misc_feature
    168776..188603
      /note="assembly_name:Contig43"
BASE COUNT 55128 a 38490 c 38018 g 54535 t 2432 others

```

## ORIGIN

```

Query Match      1.9%; Score 20; DB 76; Length 188603;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 catcaggctgggaagaga 955
|||||
Db 84037 CATACAGGCTGGGAAGAGA 84018

RESULT 33
AC023524/c
LOCUS
AC023524 195418 bp DNA HTG 20-APR-2000
Homo sapiens chromosome 4 clone RP11-301L8 map 4, WORKING DRAFT
SEQUENCE, 30 unordered pieces.
AC023524
AC023524.4 GI:7596803
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 195418)
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE
Homo sapiens chromosome 4, clone RP11-301L8
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 195418)
AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Deaellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo.A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 19, 2000 this sequence version replaced gi:7144969.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4250
Center clone name: 301_L_8
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 167354 bases at least Q40
Consensus quality: 177354 bases at least Q30
Consensus quality: 184673 bases at least Q20
Insert size: 170000; agarose-fp

```

## TITLE

## JOURNAL

## COMMENT



```

Qy 779 aatggattaatgccactatt 798
|||||
Db 173672 AATGGATTAAATGCCACTATT 173653

RESULT 34
AP000493 200000 bp DNA PRI 28-SEP-1999
LOCUS Homo sapiens genomic DNA, chromosome 3p21.3, clone:301 to 308,
DEFINITION anti-oncogene region, section 1/5.
ACCESSION AP000493
VERSION AP000493.1 GI:5926660
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Nakamura,Y., Isomura,M., Daigo,Y., Tamari,M. and Ishikawa,S.
TITLE DNA sequence analysis of a 1.2-Mb region on chromosome 3p21.3
JOURNAL Published Only in Database (1999) In press
REFERENCE 2 (bases 1 to 200000)
AUTHORS Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases. Miki
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
Japan (E-mail:mika@tokyo.jst.go.jp, Tel:81-3-5214-8491,
URL:http://www-alis.tokyo.jst.go.jp/, Fax:81-3-5214-8491,
Fax:81-3-5214-8470)
This sequence is being conducted by Japanese Foundation for Cancer
Research as a JST sequencing Team.
Principal Investigator: Yusuke Nakamura Ph.D
Phone:81-3-5449-5372, Fax:81-3-5449-5433,
yusuke@hgc.ims.u-tokyo.ac.jp
The sequence is submitted by Human Genome Sequencing in ALIS
project of JST
Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequences, please visit our
sequence archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/top.
html) or send email to webmaster@www-alis.tokyo.jst.go.jp.
FEATURES
source
1. 200000
/db_xref="taxon:9606"
/chromosome="3"
/map="3p21.3"
/complement(41441..41542)
/note="SGC33712:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4583716"
/complement(67653..67904)
/standard_name="D3S2417"
/note="CHLC.APA22D03.P17192:The location is between each
flanking site of PCR primers."
/db_xref="GDB:683838"
76856..77072
/standard_name="D3S3623"
/note="AFMB286yb1:Genethon Marker:The location is between
each flanking site of PCR primers."
/db_xref="GDB:610155"
90178..90277
/note="TIGR-A003A11:The location is between each flanking
site of PCR primers."
/db_xref="GDB:4586260"
91241..91476
/note="Cdcalc04:The location is between each flanking site
of PCR primers."
/db_xref="GDB:442017"
/complement(91341..91539)
/note="TIGR-A003B17:The location is between each flanking
site of PCR primers."
/db_xref="GDB:4586314"
/complement(91373..91547)
/standard_name="D3S4247"
/note="SHGC-16795:The location is between each flanking
site of PCR primers."
/db_xref="GDB:734134"
/complement(91438..91565)
/note="WI-18659:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4574316"
52668 a 41602 c 43186 g 62366 t 178 others
ORIGIN

Query Match 1.9%; Score 20; DB 91; Length 200000;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 778 aatggattaatgccactatt 797
|||||
Db 58801 AATGGATTAAATGCCACTAT 58820

RESULT 35
AF130343/c
LOCUS AF130343 292721 bp DNA PRI 09-JUL-1999
DEFINITION Homo sapiens chromosome 8 clone PAC 87.2 map 8q24.1, complete
sequence.
ACCESSION AF130343
VERSION AF130343.1 GI:5430715
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 292721)
AUTHORS Gloeckner,G., Rosenthal,A., Luedecke,H.-J. and Horsthemke,B.
TITLE Annotation of 642 kb of human genomic sequence in 8q24.1 Segment 2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 292721)
AUTHORS Gloeckner,G., Rosenthal,A., Luedecke,H.-J. and Horsthemke,B.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1999) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
1. 292721
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8q24.1"
/clone="PAC 87.2"
1..1333
/note="5207, LINE/L1, pos., in, rep.: 2423..4802"
/rpt_family="L1MC1"
/evidence=not_experimental
repeat_region
1..19318
/note="Incomplete at 5' end: 155005 bases on segment 1"
/misc_feature
1
/note="Region: H1437Q13"
misc_feature
1
/note="Region: end of AF130342"
1..121678
/note="Incomplete at 5' end: 22640 bases on segment 1"
/misc_feature
1
/note="Region: 10653Q13"
/complement(1334..1648)
/note="2115, SINE/Alu, pos., in, rep.: 310..0"
/rpt_family="AluYa8"
/evidence=not_experimental
repeat_region
1649..2294
/note="5207, LINE/L1, pos., in, rep.: 4802..5432"
/rpt_family="L1MC1"
/evidence=not_experimental
repeat_region
2295..2591

```

repeat\_region /note="2068, SINE/Alu, pos., in, rep.:, 9. .306"  
/rpt\_family="AluSx"  
/evidence-not\_experimental  
2592. .3640  
/note="5207, LINE/L1, pos., in, rep.:, 5432. .6326"  
/rpt\_family="L1MC1"  
/evidence-not\_experimental  
5825. .5883  
repeat\_region /note="214, LINE/L2, pos., in, rep.:, 2651. .2746"  
/rpt\_family="L2"  
/evidence-not\_experimental  
6590. .6613  
misc\_feature /note="195, (CA)n, Simple Repeat, pos., in, rep.:, 71. .94"  
/note="Region: Simple Repeats"  
/evidence-not\_experimental  
complement(7376. .7619)  
repeat\_region /note="738, LINE/L1, pos., in, rep.:, 7730. .10"  
/rpt\_family="L1MD3"  
/evidence-not\_experimental  
complement(7621. .7674)  
misc\_feature /note="214, purine-rich, Low\_complexity, pos., in, rep.:,  
55. .0"  
/note="Region: Simple Repeats"  
/evidence-not\_experimental  
complement(7688. .7868)  
repeat\_region /note="987, LINE/L1, pos., in, rep.:, 6799. .1178"  
/rpt\_family="L1MC4"  
/evidence-not\_experimental  
complement(7869. .7893)  
misc\_feature /note="204, (TAA)n, Simple Repeat, pos., in, rep.:, 96. .0"  
/note="Region: Simple Repeats"  
/evidence-not\_experimental  
complement(7894. .8062)  
repeat\_region /note="987, LINE/L1, pos., in, rep.:, 6625. .1352"  
/rpt\_family="L1MC4"  
/evidence-not\_experimental  
complement(8095. .8682)  
repeat\_region /note="1529, LINE/L1, pos., in, rep.:, 6106. .62"  
/rpt\_family="L1ME1"  
/evidence-not\_experimental  
complement(8740. .9446)  
repeat\_region /note="1374, LINE/L1, pos., in, rep.:, 3339. .2807"  
/rpt\_family="L1M4"  
/evidence-not\_experimental  
complement(9449. .10571)  
repeat\_region /note="1649, LINE/L1, pos., in, rep.:, 2343. .3803"  
/rpt\_family="L1MEC"  
/evidence-not\_experimental  
complement(10742. .11031)  
repeat\_region /note="498, LINE/L1, pos., in, rep.:, 1295. .4851"  
/rpt\_family="L1MEC"  
/evidence-not\_experimental  
1277. .12458  
repeat\_region /note="371, SINE/MIR, pos., in, rep.:, 5. .197"  
/rpt\_family="MIR"  
/evidence-not\_experimental  
12464. .12485  
repeat\_region /note="0.95, %, gtgaaccatataaattgtca, 21/22"  
/rpt\_type=Inverted  
/evidence-not\_experimental  
12514. .12535  
repeat\_region /note="0.95, %, caacttggtataactttaacggt, 21/22"  
/rpt\_type=Inverted  
/evidence-not\_experimental  
12582. .13066  
repeat\_region /note="1353, LTR/MaLR, pos., in, rep.:, 7. .464"  
/rpt\_family="LTR1C"  
/evidence-not\_experimental  
complement(13418. .13895)  
repeat\_region /note="181, LINE/L2, pos., in, rep.:, 2150. .606"  
/rpt\_family="L2"  
/evidence-not\_experimental  
complement(13492. .13536)

repeat\_region /note="100, (ORF:, 642722, 642724, (0)), GSPVHCIPPHYTECM,  
4. Type-1, copper, (blue), proteins, signature,  
[CA].[0,2][YSA].[0,1][VFY].C.[1,2][PG].[0,1]H.[2,4][MQ],  
PDOC00174, PS00196"  
/note="Region: PROSITE"  
/evidence-not\_experimental  
complement(14039. .14332)  
repeat\_region /note="494, LTR/MaLR, pos., in, rep.:, 410. .0"  
/rpt\_family="LTR1L1"  
/evidence-not\_experimental  
complement(14714. .14766)  
repeat\_region /note="184, LINE/L2, pos., in, rep.:, 213. .2543"  
/rpt\_family="L2"  
/evidence-not\_experimental  
15095. .15169  
repeat\_region /note="226, SINE/MIR, pos., in, rep.:, 49. .125"  
/rpt\_family="MIR"  
/evidence-not\_experimental  
15542. .15559  
misc\_feature /note="100, (ORF:, 0, 2, (0)), CSMCHL, 3, Cytochrome, c,  
family, heme-binding, site, signature,  
C["CPWHF"]["CPWRJCH"]["CFYFW"], PDOC00169, PS00190"  
/note="Region: PROSITE"  
/evidence-not\_experimental  
16051. .16874  
repeat\_region /note="651, LINE/L2, pos., in, rep.:, 1842. .2705"  
/rpt\_family="L2"  
/evidence-not\_experimental  
16875. .17086  
repeat\_region /note="1064, LINE/L1, pos., in, rep.:, 6012. .6210"  
/rpt\_family="L1MA5A"  
/evidence-not\_experimental  
complement(17068. .17154)  
repeat\_region /note="683, rRNA, pos., in, rep.:, 88. .33"  
/rpt\_family="5S"  
/evidence-not\_experimental  
17155. .17253  
repeat\_region /note="442, LINE/L1, pos., in, rep.:, 6197. .6295"  
/rpt\_family="L1MA5A"  
/evidence-not\_experimental  
complement(17397. .17489)  
repeat\_region /note="244, LINE/L2, pos., in, rep.:, 2705. .5"  
/rpt\_family="L2"  
/evidence-not\_experimental  
18338. .18384  
misc\_feature /note="23, AT-rich, Low\_complexity, pos., in, rep.:,  
1. .47"  
repeat\_region /note="Region: Simple Repeats"  
/evidence-not\_experimental  
19044. .19170  
repeat\_region /note="260, LTR/MaLR, pos., in, rep.:, 45. .137"  
/rpt\_family="LTR1H"  
/evidence-not\_experimental  
19880. .20225  
repeat\_region /note="1162, LTR/MaLR, pos., in, rep.:, 1. .365"  
/rpt\_family="LTR1A1"  
/evidence-not\_experimental  
20907. .21231  
repeat\_region /note="881, LTR/Retroviral, pos., in, rep.:, 113. .450"  
/rpt\_family="LTR16A"  
/evidence-not\_experimental  
complement(21327. .21384)  
repeat\_region /note="189, LINE/L2, pos., in, rep.:, 2612. .98"  
/rpt\_family="L2"  
/evidence-not\_experimental  
21385. .21687  
repeat\_region /note="1903, SINE/Alu, pos., in, rep.:, 1. .303"  
/rpt\_family="AluSx"  
/evidence-not\_experimental  
complement(21688. .21718)  
repeat\_region /note="189, LINE/L2, pos., in, rep.:, 2555. .155"  
/rpt\_family="L2"  
/evidence-not\_experimental

```

repeat_region 21796..21922
/note="236, LTR/MaLR, pos., in, rep.: 121..240"
/rpt_family="MLT1J"
/evidence=not_experimental
misc_feature 22022..22052
/note="25, AT-rich, Low_complexity, pos., in, rep.: 1..31"

Query Match 1.9%; Score 20; DB 89; Length 292721;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 948 ggaagaatgcagatgagat 967
|||||
Db 140332 GGAAGAATGCATGGAGAT 140313

RESULT 36
AP002534 300000 bp DNA PRI 17-JUN-2000
LOCUS Homo sapiens genomic DNA, chromosome 1q22-q23, CD1 region, section 3/4.
DEFINITION AP002534
ACCESSION AP002534
VERSION AP002534.1 GI:8570523
KEYWORDS Homo sapiens DNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Shiina,T., Takishima,N. and Inoko,H.
DNA sequence analysis of a 1.1-Mb region on chromosome 1q22-q23
Published Only in DataBase (2000) In press
REFERENCE 2 (bases 1 to 300000)
Hirakawa,M. and Yamaguchi,H.
Direct Submission
Submitted (15-JUN-2000) to the DDBJ/EMBL/GenBank databases. Mika
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
Japan [E-mail:mika@tokyo.jst.go.jp,
URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
Fax:81-3-5214-8470]
This sequence is conducted by Tokai University
as a JST sequencing Team.
Principal Investigator: Hidetoshi Inoko Ph.D
Phone:+81-463-93-1121, Fax:+81-463-94-8884,
The sequence is submitted by Human Genome Sequencing in ALIS
project of JST
Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequences,
please visit our sequence archive Web site
(http://www-alis.tokyo.jst.go.jp/HGS/top.html)
or send email to webmaster@www-alis.tokyo.jst.go.jp.
FEATURES
source
1..300000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q22-q23"
complement(41083..41209)
/note="PrG-11813/PrG-11814;The location is between each
flanking site of PCR primers."
/db_xref="GDB:598189"
misc_feature 56196..74273
/gene="MND4"
/note="match MND4 (NM_002432) with account of intervene
sequences."
gene 56196..74273
/gene="MND4"
STS 74039..74269
/gene="MND4"
/note="RH95925;The location is between each flanking site

```

```

of PCR primers."
/db_xref="GDB:588364"
74045..74261
/gene="MND4"
/note="RH17742;The location is between each flanking site
of PCR primers."
/db_xref="GDB:4572245"
181942..182031
/note="SHGC-8115;The location is between each flanking
site of PCR primers."
/db_xref="GDB:1232045"
complement(226219..226439)
/note="WI-6674;The location is between each flanking site
of PCR primers."
/db_xref="GDB:588221"
232156..277366
/gene="IFI16"
/note="match IFI16 (NM_005531) with account of intervene
sequences."
232156..277366
/gene="IFI16"
complement(241289..241444)
/note="RH48430;The location is between each flanking site
of PCR primers."
272547)
complement(272376..272547)
/note="RH34584;The location is between each flanking site
of PCR primers."
/db_xref="GDB:355027"
277112..277257
/gene="IFI16"
/note="RH103;The location is between each flanking site of
PCR primers."
/db_xref="GDB:4559702"
277120..277321
/gene="IFI16"
/note="RH37786;The location is between each flanking site
of PCR primers."
/db_xref="GDB:6455398"
282987..283075
/note="SHGC-18224;The location is between each flanking
site of PCR primers."
/db_xref="GDB:1230561"
complement(284696..299063)
/gene="AIM2"
complement(284696..299063)
/gene="AIM2"
/note="match AIM2 (NM_004833) with account of intervene
sequences."
BASE COUNT 94202 a 57479 c 56393 g 91926 t
ORIGIN

Query Match 1.9%; Score 20; DB 91; Length 300000;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 778 aaatgattatgccactat 797
|||||
Db 144739 AAATGGATTATGCCACTAT 144758

RESULT 37
AP001118/c 347550 bp DNA BCT 07-SEP-2000
LOCUS Buchnera sp. APS genomic DNA, complete sequence, segment 1/2.
DEFINITION AP001118 AP000398 BA000003
ACCESSION AP001118.1 GI:10038693
VERSION AP001118.1
KEYWORDS ferredoxin II; GTP-binding protein; membrane protein; kinase.
SOURCE Buchnera sp. APS (strain:APS, isolate:Tokyo1998,
specific:host:Acyrthosiphon pisum) DNA.
ORGANISM Buchnera sp. APS
Bacteria; Proteobacteria; gamma subdivision; Buchnera.
REFERENCE 1 (sites)

```

AUTHORS Shigenobu,S., Watanabe,H., Hattori,M., Sakaki,Y. and Ishikawa,H.  
TITLE Genome sequence of the endocellular bacterial symbiont of aphids  
JOURNAL Buchnera sp. APS  
MEDLINE Nature 407 (6800), 81-86 (2000)  
REFERENCE 20445173  
AUTHORS 2 (bases 1 to 347550)  
TITLE Shigenobu,S., Hattori,M., Watanabe,H., Toyoda,A., Yada,T.,  
Sakaki,Y. and Ishikawa,H.  
JOURNAL Direct Submission  
COMMENT Submitted (24-JAN-2000) to the DDBJ/EMBL/GenBank databases, Shuji  
Shigenobu, University of Tokyo, Department of Biological Sciences;  
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan  
(E-mail:shige@sc.riken.go.jp, URL:http://buchnera.gsc.riken.go.jp,  
Tel:81-3-5800-3553, Fax:81-3-5800-3553)  
Collaboration Information:  
RIKEN, Genomic Sciences Center (GSC)  
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,  
Japan  
Tel:81-42-778-9923 FAX:81-42-778-9924  
AP000398: Submitted(23-Aug-1999).  
FEATURES  
source Location/Qualifiers  
1. .347550  
/organism="Buchnera sp. APS"  
/isolate="Tokyo1998"  
/specific\_host="Acyrtosiphon pisum"  
/strain="APS"  
/db\_xref="taxon:107806"  
/note="primary endosymbiont of Acyrtosiphon pisum"  
197. .2083  
/gene="gida"  
197. .2083  
/gene="gida"  
/note="BU001"  
/codon\_start=1  
/transl\_table=11  
/product="glucose inhibited division protein A"  
/protein\_id="BAB12729.1"  
/db\_xref="GI:10038694"  
/translation="MENLRFDDVVVGAGHAGTAAMASSRGCKTLILLTQKTSDLGA  
LSCNPAIGGICKSHLVKEIDALGGMAKAIDYSGIOPRLNLSKSGPAVSTRQAADKI  
LYHETVKLLKKNLNLLEAEVKDLIFKNYSVGVLTQNEINFYSRSVYVLAAGTFLG  
KTHIGKYSAGRIDGSAIDLSVRLRELSRVNRLKTGTPPRIDINTVFNFLLIQ  
NSDTPVPFVFMGNVSHHPKQIPCYLTHNEKTHEIRKNDKSPIYTFGLKGLGPRY  
CPSTEDKIVRPDKSHQVFLPEGLSSIKVYPNGISTSLPIEVQBIQVASIKGLEKS  
KIIRPGVAIDYDFDPKDLNLTLKSLIKGLFAGQINGTGYVEAASQGLAGLAAA  
LSKNTEGFPDRQAYLGVLIIDDLFTQCTEETPYRMTSPAERYLSLRDNADRLTE  
IGKRLGVNDSRIRINQKVLNLTQTEMNRLKKNKISPISPDAILKLYNINLIKELS  
MSELLKRPQRYQDLSLESFRGIVDLEAIGQIEINEIKYAGYIKKQSEIEIRHLKNE  
NTFLSSYDYNKIRGLSSEYVKKLNDYKPISIGQASRISGITPPAAISILLHLKESY  
KHTL"  
2278. .3102  
/gene="atpB"  
2278. .3102  
/gene="atpB"  
/EC\_number="3.6.1.34"  
/note="BU002"  
/codon\_start=1  
/transl\_table=11  
/product="ATP synthase A chain"  
/protein\_id="BAB12730.1"  
/db\_xref="GI:10038695"  
/translation="MILEKISDPQKYISHHLSHLQIDLRSFKIIQPGALSDDYTVNV  
DSMFFSLVLSGFFSLPYMVGKKITQIGPKLOTALBELIFEVLNVKSMYQGNALI  
APLSLTVINVLNMLDLPIDFPFPISEKVELPAMRIVPSADINITLSMSLGVFF  
LILFYVKKIGYVGLKELLIQPNHPVFIFFNFLEFVSLVSKPISLGLRLGNMYA  
GEMIFILIAGLPMPWTCQFLNPWAIFPHILIIISLQAFIFMVLTVIYLSMASQSHKD"  
3139. .3378  
/gene="atpE"  
3139. .3378  
/gene="atpE"  
/EC\_number="3.6.1.34"  
/note="BU003"  
/codon\_start=1

/transl\_table=11  
/product="ATP synthase C chain"  
/protein\_id="BAB12731.1"  
/db\_xref="GI:10038696"  
/translation="MENLNVDMLYTAVAIMVGLASIAAIGIGILGGKFLEGAARQPD  
LVPLLRTQFFVVMGLVDPAIPMTIAVGLGYMLFAIS"  
3497. .3982  
/gene="atpF"  
3497. .3982  
/gene="atpF"  
/EC\_number="3.6.1.34"  
/note="BU004"  
/codon\_start=1  
/transl\_table=11  
/product="ATP synthase B chain"  
/protein\_id="BAB12732.1"  
/db\_xref="GI:10038697"  
/translation="MNLNATILGQAISFVLVFWFCMKYIWPPTILAIETROKEIKESL  
TNAKARDEYILEKRIHQNIIDAKOKASNILNSANKQKVSILEDARNQALEESKKII  
LNKQSEINAIATHARKNLHKEVVVDLSISMAEKIIRKNISKDDNQELDELVTLSLVK  
N"  
3982. .4515  
/gene="atpH"  
3982. .4515  
/gene="atpH"  
/EC\_number="3.6.1.34"  
/note="BU005"  
/codon\_start=1  
/transl\_table=11  
/product="ATP synthase delta chain"  
/protein\_id="BAB12733.1"  
/db\_xref="GI:10038698"  
/translation="MSVADTIARPYAQAIFEIAIENNTIEKWNILIFIKTIAHKKF  
KNFLSGISPKYLSLIFITGTNIIDENAKNLKLSSEQRNFILNIFERFVKLEAC  
YKNIIVOLKSAPSLKENLINKRVLERFLFKTKIYKVDPNILNGMIVKVNNTIF  
DLSAQNHKLQSLDSLNF"  
4530. .6068  
/gene="atpA"  
4530. .6068  
/gene="atpA"  
/EC\_number="3.6.1.34"  
/note="BU006"  
/codon\_start=1  
/transl\_table=11  
/product="ATP synthase alpha chain"  
/protein\_id="BAB12734.1"  
/db\_xref="GI:10038699"  
/translation="MRLNSTEIKLKERIAQFEVFNQSYNEGSIISVSDGIKINGL  
SNVLMGEMILLPNNEYAIALNIERDVTGAVVMGPYIHISEGAKVRCGTGKILEVPVGON  
FLGRVNALGFPIDGKDSIQNDGFFPVEADAPGVIDRKSQVNPQIQGYKVIDSMPIG  
RGORELIGDRTGKTALADITTIINQKSGIKCIYVAIGKQLSTIINNVKKLENNAL  
FNTIIVVASASEAASLQYLAPYSGCMAEFERNKEDSLIYDDLSKHAYAYRQISLL  
LRRPGREAPPGDIFVYLSRLLEASRVSMYVQKTKNKITGTSITALPIETQS  
GDVSAFVPTNVIITDGOIFELSNLNSGIRPAVNPGLISVRVGSAAOTTIIRKLSGG  
IRTALAQYBELAFAFSQSSDLDDTRKQLNHGQKIELLAKQKQYSPISIAEQALILFV  
AENNLDSDISDKITKFEILLIYAHNYIFDLMEENIKTGDFNIVIKDKFIKLITEK  
KNQF"  
6101. .6973  
/gene="atpG"  
6101. .6973  
/gene="atpG"  
/EC\_number="3.6.1.34"  
/note="BU007"  
/codon\_start=1  
/transl\_table=11  
/product="ATP synthase gamma chain"  
/protein\_id="BAB12735.1"  
/db\_xref="GI:10038700"  
/translation="MTSEIKKKNKIVSVTNTKTKTAMEMVAVSKMKRKTETMRSGRP  
YSDIIRKVIDHTVQGNLEYKHSLEERTNRIGMIIISTDRGLCGGLNLFKQVLFK  
IQNFKAVNIPCDLILFGLKSLSVFKLGSNLAKATNLGENPKLEELINSVGILOEY  
OCKRDKIFITAYNKFHNMKSQYPTITOLLPFSSKKNDQDASNNNDYLYEPESKILDT  
LFNRYIESQVYSILENIASEAARMIAKMTATDNSGNRIKELQVLQVNVKVRQANITQE

```

gene      LNEIVSGASAVSID"
          6997..8394
          /gene="atpD"
CDS       6997..8394
          /gene="atpD"
          /EC_number="3.6.1.34"
          /note="BU008"
          /codon_start=1
          /transl_table=11
          /product="ATP synthase beta chain"
          /protein_id="BAB12736.1"
          /db_xref="GI:10038701"
          /translation="MATKIILIIIGAVVDVEFNQSPKIVYNALEVNKNQYKLIILEVQ
          QOLGSGVVTAMGSSNGKLRLGLIIVTKPVGATLGRILNVLGETIDNKGAL
          KNNQSDKTEYIEHPSKPYRDQASCREILETGKIVIDLICPFSKGGKVLFGGAGVG
          KTVNMELIRINAVHSGYSYVTGVTREGNDEYHEMNDQSVQLDKVSLVYQGNNEP
          PGNRLURVATGLTIAERDEGKDWLLFIDNIRYTLAGTEVSALLGRMPSAVGQPT
          LAEMGLQERTITKNGSITSVQAVYPADLLTDPSPATFAHLDSTVTSRQIASL
          GIIYPAIDPLNSTSRQLDPYIIVGDEHYETALGVQSILQRYQELKDIILAILGMDLAERD
          KLLVSRARKIQRFLSQPFFVAEVFTGFPKIVSLKDNIRAFKGIILKGEFFDLDLPEQAFY
          MVGSTEEVIERAKKL"
          8421..8837
          /gene="atpC"
          /EC_number="3.6.1.34"
          /note="BU009"
          /codon_start=1
          /transl_table=11
          /product="ATP synthase epsilon chain"
          /protein_id="BAB12737.1"
          /db_xref="GI:10038702"
          /translation="MNFYLDVVSLLTKTIFSGFVKIRVSGSGELGIYPGHAQLLSIL
          KPGMVYIFHKDKKECIYISGILEVQPSVYILADVAIHAIDLRSRLTKYKNNAE
          ESIKNNKINKADILQLISKEIKRLVLEVDKFK"
          complement(8911..11322)
          /gene="gyrB"
          complement(8911..11322)
          /gene="gyrB"
          /EC_number="5.99.1.3"
          /note="BU010"
          /codon_start=1
          /transl_table=11
          /product="DNA gyrase subunit B"

Query Match      1.9%; Score 20; DB 2; Length 347550;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 777 aaaaatgattaatgccacta 796
      |||||
Db 244399 AAAATGATTAAATGCCACTA 244380

RESULT 38
CNS06WMC      837 bp      DNA      STS      11-JAN-2001
LOCUS      T7 end of clone AX0AA036B03 of library AX0AA from strain CBS 7064
DEFINITION      of Pichia farinosa, sequence tagged site.
ACCESSION      AL418978
VERSION      AL418978.1 GI:12202156
KEYWORDS      STS.
SOURCE      Pichia farinosa.
ORGANISM      Pichia farinosa
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE      1 (bases 1 to 837)
AUTHORS      de Montigny,J., Spohner,C., Souciet,J., Tekala,F., Dujon,B.,
Wincker,P., Artiguenave,F. and Potier,S.
TITLE      Genomic Exploration of the Hemiascomycetous Yeasts: 15. Pichia
sorbitophila
JOURNAL      FEBS Lett. 487 (1), 87-90 (2000)
PUBLISHED      11152890

gene      LNEIVSGASAVSID"
          6997..8394
          /gene="atpD"
CDS       6997..8394
          /gene="atpD"
          /EC_number="3.6.1.34"
          /note="BU008"
          /codon_start=1
          /transl_table=11
          /product="ATP synthase beta chain"
          /protein_id="BAB12736.1"
          /db_xref="GI:10038701"
          /translation="MATKIILIIIGAVVDVEFNQSPKIVYNALEVNKNQYKLIILEVQ
          QOLGSGVVTAMGSSNGKLRLGLIIVTKPVGATLGRILNVLGETIDNKGAL
          KNNQSDKTEYIEHPSKPYRDQASCREILETGKIVIDLICPFSKGGKVLFGGAGVG
          KTVNMELIRINAVHSGYSYVTGVTREGNDEYHEMNDQSVQLDKVSLVYQGNNEP
          PGNRLURVATGLTIAERDEGKDWLLFIDNIRYTLAGTEVSALLGRMPSAVGQPT
          LAEMGLQERTITKNGSITSVQAVYPADLLTDPSPATFAHLDSTVTSRQIASL
          GIIYPAIDPLNSTSRQLDPYIIVGDEHYETALGVQSILQRYQELKDIILAILGMDLAERD
          KLLVSRARKIQRFLSQPFFVAEVFTGFPKIVSLKDNIRAFKGIILKGEFFDLDLPEQAFY
          MVGSTEEVIERAKKL"
          8421..8837
          /gene="atpC"
          /EC_number="3.6.1.34"
          /note="BU009"
          /codon_start=1
          /transl_table=11
          /product="ATP synthase epsilon chain"
          /protein_id="BAB12737.1"
          /db_xref="GI:10038702"
          /translation="MNFYLDVVSLLTKTIFSGFVKIRVSGSGELGIYPGHAQLLSIL
          KPGMVYIFHKDKKECIYISGILEVQPSVYILADVAIHAIDLRSRLTKYKNNAE
          ESIKNNKINKADILQLISKEIKRLVLEVDKFK"
          complement(8911..11322)
          /gene="gyrB"
          complement(8911..11322)
          /gene="gyrB"
          /EC_number="5.99.1.3"
          /note="BU010"
          /codon_start=1
          /transl_table=11
          /product="DNA gyrase subunit B"

Query Match      1.9%; Score 20; DB 2; Length 347550;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 777 aaaaatgattaatgccacta 796
      |||||
Db 244399 AAAATGATTAAATGCCACTA 244380

RESULT 38
CNS06WMC      837 bp      DNA      STS      11-JAN-2001
LOCUS      T7 end of clone AX0AA036B03 of library AX0AA from strain CBS 7064
DEFINITION      of Pichia farinosa, sequence tagged site.
ACCESSION      AL418978
VERSION      AL418978.1 GI:12202156
KEYWORDS      STS.
SOURCE      Pichia farinosa.
ORGANISM      Pichia farinosa
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE      1 (bases 1 to 837)
AUTHORS      de Montigny,J., Spohner,C., Souciet,J., Tekala,F., Dujon,B.,
Wincker,P., Artiguenave,F. and Potier,S.
TITLE      Genomic Exploration of the Hemiascomycetous Yeasts: 15. Pichia
sorbitophila
JOURNAL      FEBS Lett. 487 (1), 87-90 (2000)
PUBLISHED      11152890

REFERENCE 2 (bases 1 to 837)
AUTHORS      Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Liorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
TITLE      Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
yeast species for molecular evolution studies(1)
JOURNAL      FEBS Lett. 487 (1), 3-12 (2000)
PUBLISHED      11152876
REFERENCE 3 (bases 1 to 837)
AUTHORS      Genoscope.
TITLE      Direct Submision
JOURNAL      Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT      This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
    Location/Qualifiers
        1..837
            /organism="Pichia farinosa"
            /strain="CBS 7064"
            /db_xref="taxon:4920"
            /clone="AX0AA036B03"
            /clone_lib="AX0AA"
            /note="end : T7"
        misc_feature
            complement(<3..>293)
            /note="similar to Saccharomyces cerevisiae ORF YLR009w [
            similarity to ribosomal protein L24.e.B ]"
            /evidence=not-experimental
        BASE COUNT      212 a 199 c 182 g 240 t 4 others
        ORIGIN
            Query Match      1.8%; Score 19; DB 53; Length 837;
            Best Local Similarity 100.0%; Pred. No. 28;
            Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 tggtaaatctctctctgga 750
      |||||
Db 397 TGGTAAATCTCTTCCTGGA 415

RESULT 39
CNS06UMH      888 bp      DNA      STS      11-JAN-2001
LOCUS      T3 end of clone AX0AA011A10 of library AX0AA from strain CBS 7064
DEFINITION      of Pichia farinosa, sequence tagged site.
ACCESSION      AL416031
VERSION      AL416031.1 GI:12195971
KEYWORDS      STS.
SOURCE      Pichia farinosa.
ORGANISM      Pichia farinosa
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE      1 (bases 1 to 888)
AUTHORS      de Montigny,J., Spohner,C., Souciet,J., Tekala,F., Dujon,B.,
Wincker,P., Artiguenave,F. and Potier,S.
TITLE      Genomic Exploration of the Hemiascomycetous Yeasts: 15. Pichia
sorbitophila
JOURNAL      FEBS Lett. 487 (1), 87-90 (2000)
PUBLISHED      11152890
REFERENCE 2 (bases 1 to 888)
AUTHORS      Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,

```

Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,  
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,  
 Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
 Wincker,P. and Weissenbach,J.  
 Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of  
 yeast species for molecular evolution studies(1)  
 FEBS Lett. 487 (1), 3-12 (2000)  
 11152876  
 3 (bases 1 to 888)  
 Genoscope.  
 Direct Submission  
 Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,  
 2 rue Gaston Cremlieux, Cp 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)  
 This STS is part of a random genomic sequencing program of thirteen  
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.

# FEATURES

source  
 1. .888  
 /organism="Pichia farinosa"  
 /strain="CBS 7064"  
 /db\_xref="taxon:4920"  
 /clone="AX0AA011A10"  
 /clone\_lib="AX0AA"  
 /note="end : T3"  
 complement(<3. >292)  
 /note="similar to Saccharomyces cerevisiae ORF YLR009w [  
 similarity to ribosomal protein L24.e.B ]  
 1 putative frameshift(s)"  
 /evidence=not\_experimental  
 219 a 212 c 195 g 256 t 6 others

misc\_feature  
 complement(<3. >292)  
 /note="similar to Saccharomyces cerevisiae ORF YLR009w [  
 similarity to ribosomal protein L24.e.B ]  
 1 putative frameshift(s)"  
 /evidence=not\_experimental  
 219 a 212 c 195 g 256 t 6 others

BASE COUNT 219 a 212 c 195 g 256 t 6 others  
 ORIGIN  
 Query Match 1.8%; Score 19; DB 53; Length 888;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 tggtaaatctctctctgga 750  
 |||||  
 Db 396 TGGTAAATCTCTTCTCGGA 414

RESULT 40  
 CNS060PZ 891 bp DNA STS 11-JAN-2001  
 LOCUS T3 end of clone AX0AA012B09 of library AX0AA from strain CBS 7064  
 DEFINITION of Pichia farinosa, sequence tagged site.  
 ACCESSION AL416157  
 VERSION AL416157.1 GI:12196274  
 KEYWORDS STS.  
 SOURCE Pichia farinosa.  
 ORGANISM Pichia farinosa.  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Pichia.

REFERENCE 1 (bases 1 to 891)  
 de Montigny,J., Spehner,C., Souciet,J., Tekala,F., Dujon,B.,  
 Wincker,P., Artiguenave,F. and Potier,S.  
 Genomic Exploration of the Hemiascomycetous Yeasts: 15. Pichia  
 sorbitophila  
 FEBS Lett. 487 (1), 87-90 (2000)  
 11152890

REFERENCE 2 (bases 1 to 891)  
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,  
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,  
 Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
 Wincker,P. and Weissenbach,J.

## TITLE

JOURNAL  
 PUBMED  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

## COMMENT

Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,  
 2 rue Gaston Cremlieux, Cp 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)  
 This STS is part of a random genomic sequencing program of thirteen  
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.

## FEATURES

source  
 1. .891  
 /organism="Pichia farinosa"  
 /strain="CBS 7064"  
 /db\_xref="taxon:4920"  
 /clone="AX0AA012B09"  
 /clone\_lib="AX0AA"  
 /note="end : T3"  
 complement(<3. >290)  
 /note="similar to Saccharomyces cerevisiae ORF YLR009w [  
 similarity to ribosomal protein L24.e.B ]"  
 /evidence=not\_experimental  
 222 a 212 c 199 g 255 t 3 others

misc\_feature  
 complement(<3. >290)  
 /note="similar to Saccharomyces cerevisiae ORF YLR009w [  
 similarity to ribosomal protein L24.e.B ]"  
 /evidence=not\_experimental  
 222 a 212 c 199 g 255 t 3 others

Query Match 1.8%; Score 19; DB 53; Length 891;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 732 tggtaaatctctctctgga 750  
 |||||  
 Db 397 TGGTAAATCTCTTCTCGGA 415

RESULT 41  
 CNS060U6 904 bp DNA STS 11-JAN-2001  
 LOCUS T3 end of clone AX0AA012G08 of library AX0AA from strain CBS 7064  
 DEFINITION of Pichia farinosa, sequence tagged site.  
 ACCESSION AL416236  
 VERSION AL416236.1 GI:12196442  
 KEYWORDS STS.  
 SOURCE Pichia farinosa.  
 ORGANISM Pichia farinosa.  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Pichia.

REFERENCE 1 (bases 1 to 904)  
 de Montigny,J., Spehner,C., Souciet,J., Tekala,F., Dujon,B.,  
 Wincker,P., Artiguenave,F. and Potier,S.  
 Genomic Exploration of the Hemiascomycetous Yeasts: 15. Pichia  
 sorbitophila  
 FEBS Lett. 487 (1), 87-90 (2000)  
 11152890

REFERENCE 2 (bases 1 to 904)  
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,  
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,  
 Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
 Wincker,P. and Weissenbach,J.



Saurin,W., Tekaiia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.  
Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of yeast species for molecular evolution studies(1)  
FEBS Lett. 487 (1), 3-12 (2000)  
11152876  
3 (bases 1 to 904)  
Genoscope.  
Direct Submission  
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
This STS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES  
source  
Location/Qualifiers  
1..904  
/organism="Pichia farinosa"  
/strain="CBS 7064"  
/db\_xref="taxon:4920"  
/clone\_lib="AX0A012G08"  
/clone\_lib="AX0AA"  
/note="end : T3"  
complement(<3..>293)  
/note="similar to Saccharomyces cerevisiae ORF YLR009w [ similarity to ribosomal protein L24.e.B ]"  
/evidence=not\_experimental

BASE COUNT 221 a 216 c 180 g 265 t 22 others

ORIGIN  
Query Match 1.8%; Score 19; DB 53; Length 904;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 732 tggtaaatctctctctgga 750  
|||||  
Db 397 TGGTAATCTCTCTCGGA 415  
|||||

RESULT 42  
CNS06XUJ/c 1003 bp DNA STS 11-JAN-2001  
LOCUS  
DEFINITION  
T3 end of clone AY0AA003H09 of library AY0AA from strain CBS 6340 of Kluyveromyces thermotolerans, sequence tagged site.  
ACCESSION  
AL420209.1 GI:12203392  
VERSION  
AL420209.1 GI:12203392  
SOURCE  
Kluyveromyces thermotolerans.  
ORGANISM  
Kluyveromyces thermotolerans  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
REFERENCE  
1 (bases 1 to 1003)  
AUTHORS  
Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P. and Dujon,B.  
TITLE  
Genomic Exploration of the Hemiascomycetous Yeasts: 10.  
Kluyveromyces thermotolerans  
FEBS Lett. 487 (1), 61-65 (2000)  
11152885  
PUBLISHED  
2 (bases 1 to 1003)  
AUTHORS  
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaiia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.

Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of yeast species for molecular evolution studies(1)  
FEBS Lett. 487 (1), 3-12 (2000)  
11152876  
3 (bases 1 to 1003)  
Genoscope.  
Direct Submission  
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
This STS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES  
source  
Location/Qualifiers  
1..1003  
/organism="Kluyveromyces thermotolerans"  
/strain="CBS 6340"  
/db\_xref="taxon:4916"  
/clone\_lib="AY0AA003H09"  
/clone\_lib="AY0AA"  
/note="end : T3"  
/note="end : T3"  
complement(<3..>548)  
/note="similar to Saccharomyces cerevisiae ORF YOR019w [ similarity to YDR474c ]"  
/evidence=not\_experimental  
/note="end : T3"  
/note="similar to Saccharomyces cerevisiae ORF YDR474c [ similarity to C-terminal region of YOR019w ]"  
/evidence=not\_experimental

BASE COUNT 301 a 222 c 232 g 245 t 3 others

ORIGIN  
Query Match 1.8%; Score 19; DB 53; Length 1003;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 25 acgctgctctctctgctg 43  
|||||  
Db 668 ACGCTGCTCTCTGCTGG 650  
|||||

RESULT 43  
CNS06XUI 1064 bp DNA STS 11-JAN-2001  
LOCUS  
DEFINITION  
T7 end of clone AY0AA003H09 of library AY0AA from strain CBS 6340 of Kluyveromyces thermotolerans, sequence tagged site.  
ACCESSION  
AL420208.1 GI:12203391  
VERSION  
AL420208.1 GI:12203391  
SOURCE  
Kluyveromyces thermotolerans.  
ORGANISM  
Kluyveromyces thermotolerans  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
REFERENCE  
1 (bases 1 to 1064)  
AUTHORS  
Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P. and Dujon,B.  
TITLE  
Genomic Exploration of the Hemiascomycetous Yeasts: 10.  
Kluyveromyces thermotolerans  
FEBS Lett. 487 (1), 61-65 (2000)  
11152885  
PUBLISHED  
2 (bases 1 to 1064)  
AUTHORS  
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaiia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.

Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
 Wincker,P. and Weissbach,J.  
 Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of  
 yeast species for molecular evolution studies(1)  
 FEBS Lett. 487 (1), 3-12 (2000)  
 11152876  
 3 (bases 1 to 1064)  
 Genoscope.  
 Direct Submission  
 Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,  
 2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 This STS is part of a random genomic sequencing program of thirteen  
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.  
 Location/Qualifiers  
 1. .1064  
 /organism="Kluyveromyces thermotolerans"  
 /strain="CBS 6340"  
 /db\_xref="taxon:4916"  
 /clone="AY0A003H09"  
 /clone\_11b="AY0AA"  
 /note="end : T7"  
 /note="similar to Saccharomyces cerevisiae ORF YDR473c [  
 PRP3 : essential splicing factor ]"  
 /evidence=not-experimental  
 BASE COUNT 258 a 253 c 248 g 299 t 6 others  
 ORIGIN  
 misc\_feature  
 complement(<27..>272)  
 /note="similar to Saccharomyces cerevisiae ORF YDR473c [  
 PRP3 : essential splicing factor ]"  
 /evidence=not-experimental  
 BASE COUNT 258 a 253 c 248 g 299 t 6 others  
 ORIGIN  
 Query Match 1.8%; Score 19; DB 53; Length 1064;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 25 acgtgctcctctgctggg 43  
 |||||  
 Db 592 ACGTGCCTCTCTGCTGG 610  
 RESULT 44  
 AF294795 2655 bp DNA INV 14-NOV-2000  
 LOCUS Trypanosoma cruzi host cell signaling surface protein gene, partial  
 cds  
 DEFINITION  
 ACCESSION AF294795  
 VERSION AF294795.1 GI:11141618  
 KEYWORDS  
 SOURCE Trypanosoma cruzi.  
 ORGANISM Trypanosoma cruzi  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma; Schizotrypanum.  
 REFERENCE 1 (bases 1 to 2655)  
 AUTHORS Villalta,F., Pratap,S., Nde,P.N., Burns,J.M. Jr. and Lima,M.F.  
 TITLE Host cell signaling-Trypanosoma cruzi surface protein  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2655)  
 AUTHORS Villalta,F., Pratap,S., Nde,P.N., Burns,J.M. Jr. and Lima,M.F.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-AUG-2000) Department of Microbiology, Meharry Medical  
 College, 1005 D.B. Todd Jr. Blvd., Nashville, TN 37208, USA  
 FEATURES  
 source  
 1. .2655  
 /organism="Trypanosoma cruzi"  
 /db\_xref="taxon:5693"  
 <1. .>1775  
 mRNA

/product="host cell signaling surface protein"  
 <1. .>1775  
 CDS  
 /codon\_start=3  
 /product="host cell signaling surface protein"  
 /protein\_id="AAG32026.1"  
 /db\_xref="GI:11141619"  
 /translation="ARGLLVAVGNVSDGSSGKKIYKDTSVIPWTEFEKOHKSLTG  
 LIGGGSGVNIINDGTILFPVETGKKDKGAVSLIIYSSDIASWTLSKMSADGCSDPV  
 AWRREGKMMVTLPAAHRRKRGKREHLWLTNTHTIYDIGSVGDDDDAAASSLLYK  
 QRRFNRDYMVLTLPAAHRRKRGKREHLWLTNTHTIYDIGSVGDDDDAAASSLLYK  
 NAGSGHNNEELIALYKKKGDEPSLVVSVRLTAQLKRVKEVLTATWKKVDERVSKLC  
 LFSRAKNPSAGCSAVKATDGLVGLSGKSGNTWMDYELGVDAIVRKGNGVATG  
 YADGYTFHGAWEWPGVROGQNLHFANHFTLVATLSVHGEPKGDTPILMGAPMN  
 DNKTVLLGSLDKEGKWVLCGGRTTKELRNWEPETHQVAIVLRNCTQGSVYDG  
 QRVGDAICSLRNADSGISHFYIGGGGAGSGEDVPVATFATFLIYNRLDNEIRVL  
 NAKTISIPKLTDLTLAAGATGVGTARDGGAHGGSTVCGDGLPLLLLLLGLMGIAASX  
 RSWMPRRGAPQFPPTTDAHTPGSG"  
 BASE COUNT 583 a 621 c 813 g 637 t 1 others  
 ORIGIN  
 Query Match 1.8%; Score 19; DB 5; Length 2655;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 26 cgctgctcctctgctggg 44  
 |||||  
 Db 1657 CGCTGCTCTCTGCTGGG 1675  
 RESULT 45  
 AF035151/c 3564 bp mRNA ROD 06-JUL-1998  
 LOCUS Rattus norvegicus regulator of G-protein signalling 12 PDZ-less  
 variant (RGS12) mRNA, complete cds.  
 DEFINITION  
 ACCESSION AF035151  
 VERSION AF035151.1 GI:3290013  
 KEYWORDS  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE 1 (bases 1 to 3564)  
 AUTHORS Snow,B.E., Hall,R.A., Krumins,A.M., Brothers,G.M., Bouchard,D.,  
 Brothers,C.A., Chung,S., Mangion,J., Gilman,A.G., Lefkowitz,R.J.  
 and Siderovski,D.P.  
 TITLE GTPase activating specificity of RGS12 and binding specificity of  
 an alternatively spliced PDZ (PSD-95/Dlg/zo-1) domain  
 J. Biol. Chem. 273 (28), 17749-17755 (1998)  
 JOURNAL 98316347  
 MEDLINE  
 REFERENCE 2 (bases 1 to 3564)  
 AUTHORS Snow,B.E. and Siderovski,D.P.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-NOV-1997) Quantitative Biology, Amgen Institute, 620  
 University Avenue, Suite 706, Toronto, ON M5G 2C1, Canada  
 FEATURES  
 Location/Qualifiers  
 1. .3564  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /tissue\_type="brain (striatum)"  
 1. .3564  
 /gene="RGS12"  
 277..2496  
 /gene="RGS12"  
 /note="regulator of G-alpha GTPase activity; Regulator of  
 G-protein signalling-12"  
 /codon\_start=1  
 /product="RGS12 PDZ-less variant"  
 /protein\_id="AAC40154.1"  
 /db\_xref="GI:3290014"  
 /translation="MNLEKGLSDSDVFDIOOSATVSDGELTGADLKDCINNSLSSN  
 ASLPSVQSCRRLRRERVASNAVSFRLLDQPGVGYPSDFLRKFESENIILFWQACEC

FSHVPAHDKKELSYRAREIFSKFLCSKATTPVNIQDQAQADDDILNAPHPDMFKEQOL  
QIFNLKFSYTRFLKSQLYQECVLAIEVEGRTPDSQQVPSPPASKHSISSDHSNVST  
PKLGGKSGRSLNEDVGEDESEKRGAFPSWRSRSTGRSOKKKDHGDHDAHDLH  
ANGGLCRRESQGSVSSAGSLDLSEACRTSALERDKAAKCCVHLPDGATSCVVAKSGF  
SIKEILSGLCERHGINGAAYDLFLVGGDKPLVLHODSSILATRDRLRGKRTLFRDLV  
PINRSVGLKAKPTKPVTEVLRPVVARYGLDGLSLVRLSGEKEPLDLGAPISSLDGQR  
VILEERDPGRKVTQKQKGAQVQKSSAVNNSPRNHSAMGEERTLGKNSIKIRGENG  
KSARDPRLSKREESIAKIGKKYQKINLDEAEFFELISKAQSNRADDQORGLLRKEDL  
VLPEFLRPLPPSSSELALSPPPVKGFSGKRAVTSHGEGAVQTEESYSDSPATSPASAO  
SPCSAYSPGSAHSPGSAHSPGSAHSTPGPGTAQPGKPTKPCISTVQEGTTQAWRR  
LSPELEAGGIQVVEEQVADLTLMGEGDISSPNSLPPPLPQDTPGTPRPGTSRF

BASE COUNT 873 a 941 c 963 g 787 t  
ORIGIN

Query Match 1.8%; Score 19; DB 94; Length 3564;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ggtggccagatgctgctg 106  
|||||  
DB 1404 GGTGGCAGGATGCTGCTG 1386

Search completed: November 22, 2001, 02:22:03  
Job time: 7242 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2001, 01:03:51 ; Search time 129.83 Seconds  
(without alignments)  
5034.627 Million cell updates/sec

Title: US-09-854-844-1  
Perfect score: 1041  
Sequence: 1 atgggcctctgctgtgc.....ttattacaattgaaatga 1041

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	1.9	2226	19 AAV11704	Enterocin-900 oper
2	20	1.9	2226	20 AAX17728	Bacteriocin entero
3	18	1.7	215	20 AAZ42155	Human normal bladd
4	18	1.7	699	18 AAT79127	Human serine prote
5	18	1.7	750	20 AAZ42170	Human normal bladd
6	18	1.7	1034	21 AAA61559	cDNA encoding mous
7	18	1.7	1232	22 AAF24159	Human secreted pro
8	18	1.7	1268	22 AAF24181	Human secreted pro
9	18	1.7	1268	22 AAF24182	Human secreted pro
10	18	1.7	1305	22 RAD02556	Human seripancrin
11	18	1.7	1413	21 AAD01122	Human orphan G pro

12	18	1.7	1413	21 AAA46021	Human G protein co
13	18	1.7	1413	21 AAA15587	Human G-protein co
14	18	1.7	1479	22 AAD02557	Human seripancrin
15	18	1.7	1959	21 AAA07582	Marigold beta-cycl
16	18	1.7	2038	20 AAX87154	Human protease HUP
17	18	1.7	2063	21 AAA37099	Human PRO1570 (UNQ
18	18	1.7	2063	22 AAF92113	Human PRO1570 cDNA
19	18	1.7	2063	22 AAF54396	DNA encoding prote
20	18	1.7	2070	21 AAZ90471	Cancer specific ge
21	18	1.7	2100	21 AAZ15586	Human G-protein co
22	18	1.7	3540	22 AAF26183	Human G-protein co
23	18	1.7	4091	17 AAT39809	Mouse H74 gene. M
24	18	1.7	5035	21 AAC75782	Human OREF ORF1337
25	18	1.7	28866	20 AAX22304	Human IL-1ra BAC c
26	18	1.7	160552	22 AAD02697	Human glycosyl sul
27	17	1.6	300	21 AAA00012	Human colon cancer
28	17	1.6	432	21 AAZ98178	Human signal pepti
29	17	1.6	451	22 AAF17622	Human breast cance
30	17	1.6	468	21 AAA73713	5' extension of PI
31	17	1.6	660	21 AAA73714	Human PITSLRE prot
32	17	1.6	677	20 AAX40117	Gastric cancer ass
33	17	1.6	701	20 AAX23298	Human TRY1 trypsin
34	17	1.6	725	20 AAX78022	Chimeric serine pr
35	17	1.6	741	9 AAN81633	Human spleen tryps
36	17	1.6	744	16 AAT03999	Human pancreatic t
37	17	1.6	744	16 AAT04000	Human pancreatic t
38	17	1.6	744	16 AAT04001	Human pancreatic t
39	17	1.6	771	21 AAF15875	Human prostate can
40	17	1.6	790	19 AAV24548	Trypsinogen-like p
41	17	1.6	807	17 AAT15535	DNA encoding rat c
42	17	1.6	954	20 AAZ77512	Human ovarian tumo
43	17	1.6	1045	19 AAV63911	Helicobacter pylor
44	17	1.6	1102	19 AAX14433	H. pylori GPO 783
45	17	1.6	1263	20 AAZ20365	Gene encoding bact

ALIGNMENTS

RESULT 1  
AAV11704  
ID AAV11704 standard; DNA; 2226 BP.  
XX  
AC AAV11704;  
XX  
DT 14-SEP-1998 (first entry)  
XX  
DE Enterocin-900 operon.

XX Enterocin-900; bacteriocin; signal peptide; protein secretion;  
KW preservative; food spoilage; lactic acid bacterium; antibacterial;  
KW ss.  
XX Enterococcus faecium 900.  
XX  
FH Key Location/Qualifiers  
FT CDS 1256..1471  
FT /tag= a  
FT sig\_peptide 1256..1309  
FT /\*tag= b  
FT mat\_peptide 1310..1468  
FT /\*tag= c  
XX  
XX WO9809639-A1.  
PN  
XX  
PD 12-MAR-1998.  
XX  
XX  
PF 05-SEP-1997; 97WO-US15609.  
XX  
PR 05-SEP-1996; 96US-0026257.  
XX  
PA (UYAL-) UNIV ALBERTA.  
XX

PI Franz CM, Greer GG, Leisner JJ, McCormick JK, McMullen LM;  
 PI Pkns J, Poona, Stiles ME, Van Belkum MJ, Vederas JC;  
 PI Worobo RJ, Worobo RW;  
 XX  
 DR WPI; 1998-193319/17.  
 DR P-PSDB; AAW59024.

XX Bacterial growth inhibiting peptide(s) enterocin 900 or brochohin C  
 PT - used to inhibit growth of susceptible bacteria in e.g. foodstuff,  
 PT living animal, food preparation area or fermentation vessel  
 XX  
 PS Claim 1; Page 136-137; 174pp; English.

XX This DNA sequence comprises the enterocin-900 operon of  
 CC Enterococcus faecium 900, including an open reading frame for  
 CC enterocin-900 (see AAW59024). Enterocin-900 is a novel broad  
 CC spectrum bacteriocin that has activity against other strains of  
 CC Enterococcus as well as many other organisms. The invention  
 CC includes a method for inhibiting pathogenic bacteria by providing  
 CC a bacteriocin selected from enterocin-900 and brochohin-C (see  
 CC AAW59024), either as a composition or by providing a bacterial  
 CC source of the bacteriocins. This is used to inhibit spoilage  
 CC bacteria in foodstuffs, such as meat, inhibit pathogenic bacteria  
 CC topically in animals, e.g. to treat mastitis (claimed), and inhibit  
 CC bacterial infection of fermentation reactors.

XX Sequence 2226 BP; 756 A; 393 C; 301 G; 776 T; 0 other;  
 SQ

Query Match 1.9%; Score 20; DB 19; Length 2226;  
 Best Local Similarity 100.0%; Pred. No. 2.7;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 819 cttctctgactcttctgttc 838  
 Db 547 cttctctgactcttctgttc 566  
 |||||

RESULT 2  
 AAX17728  
 ID AAX17728 standard; DNA; 2226 BP.

AC AAX17728;  
 XX  
 XX 25-MAY-1999 (first entry)  
 DT  
 XX Bacteriocin enterocin 900 forward operon.

XX Enterocin 900; brochohin-C; bacterial growth; microorganism; inhibit;  
 KW food; meat; fermentation; bacteriocin; mastitis; secretory vector; lcaC;  
 KW leucocin A; lcaD; enzyme; probiotic; L. gelidium; Listeria; Lactobacillus;  
 KW sulphide-producing; spoilage; sensory property; smell; taste; infection;  
 KW ground beef; silage; Salmonella; poultry; intestine; therapeutic; colour;  
 KW mucosal tissue; vacuum-packed; storage; divergicin; immunity peptide; ss.

XX Enterococcus faecium.

XX WO9902555-A1.

XX 21-JAN-1999.

XX 09-JUL-1998; 98WO-US14331.

XX 05-SEP-1997; 97US-0924629.

XX 09-JUL-1997; 97US-0052072.

XX (UYAL-) UNIV ALBERTA.

XX Franz C, Greer GG, Leisner JJ, McCormick JK, McMullen LM;  
 PI Poon A, Stiles ME, Van Belkum MJ, Vederas JC, Worobo RJ;  
 PI Worobo RW;

XX WPI; 1999-120779/10.

XX New bacteriocins enterocin 900 and brochohin-C - useful as food  
 PT preservatives and therapeutic antimicrobials  
 XX  
 PS Claim 1; Page 135-136; 182pp; English.

XX The invention relates to peptides enterocin 900, brochohin-C that inhibit  
 CC bacterial growth. Microorganisms expressing the above peptides are used  
 CC to inhibit bacterial growth in foods (specifically meat), live animals  
 CC (applied topically), food preparation areas and fermentation vessels. The  
 CC bacteriocins enterocin 900, brochohin-C may be used directly, specifically  
 CC to inhibit bacteria that cause mastitis. Antibodies specific for the  
 CC peptides are used to detect expression of bacteriocins in cells. Secretory  
 CC vectors (containing sequences for a promoter, a leucocin A processing  
 CC peptide, a heterologous peptide, lcaC and lcaD) are used to secrete  
 CC bacteriocins, enzymes or other proteins, e.g. for use in food production  
 CC and as probiotics. L. gelidium, or other leucocin A-expressing bacteria,  
 CC inhibit the growth of Listeria and the sulphide-producing spoilage  
 CC organism Lactobacillus sake; they also improve the sensory properties  
 CC (smell, taste and colour) of meat, specifically ground beef. L. gelidium  
 CC can also be used in preparation of animal feeds (silage), as probiotic,  
 CC to control Salmonella in poultry intestines and therapeutically against  
 CC mucosal tissue infections. Treatment with bacteriocins provides vacuum-  
 CC packed meat products with predictable and longer storage life. The  
 CC secretion vectors can express several bacteriocins for broader spectrum  
 CC of activity, e.g. tailored for a particular target organism.

XX Sequence 2226 BP; 756 A; 393 C; 301 G; 776 T; 0 other;  
 SQ

Query Match 1.9%; Score 20; DB 20; Length 2226;  
 Best Local Similarity 100.0%; Pred. No. 2.7;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 819 cttctctgactcttctgttc 838  
 Db 547 cttctctgactcttctgttc 566  
 |||||

RESULT 3

AAX42155  
 ID AAX42155 standard; cDNA; 215 BP.

AC AAX42155;

XX 31-JAN-2000 (first entry)

XX Human normal bladder tissue cDNA derived EST 34.

XX Human; bladder; treatment; EST; expressed sequence tag; cytostatic;  
 KW cancer; gene therapy; ss.

XX Homo sapiens.

XX DE19818620-A1.

XX 28-OCT-1999.

XX 21-APR-1998; 98DE-1018620.

XX 21-APR-1998; 98DE-1018620.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

XX WPI; 1999-602416/52.

XX P-PSDB; AAY60359, AAY60360, AAY60361.

XX New polypeptides and their nucleic acids, useful for treatment of  
 PT bladder tumour and identification of therapeutic agents -

XX Claim 3; Page 174; 366pp; German.

XX This invention describes novel polypeptide fragment sequences (I) and  
 CC their encoding nucleic acids (II) which are highly expressed in normal  
 CC bladder tissue and have cytostatic activity. (II) are used for  
 CC recombinant expression of (I) and to isolate complete genes. (I) are  
 CC used to identify agents suitable for the treatment of bladder tumours,  
 CC to directly treat this form of cancer (including expression from gene  
 CC therapy vectors), or are used in a preparation for cancer treatment. (I)  
 CC is also used for the generation of specific antibodies. (II) are  
 CC identified by assembling ESTs (expressed sequence tags) from a  
 CC particular tissue type before comparison of expression patterns. This  
 CC allows a significantly longer fragment of the gene to be revealed, and  
 CC therefore reduces the number of failures because of ESTs from different  
 CC libraries representing different parts of the same unknown gene  
 CC distorting the estimated frequency of occurrence in a particular tissue.  
 CC AA242122-242248 represent EST fragments derived from a human normal  
 CC bladder tissue cDNA library which encode the protein fragments  
 CC represented in AAY60329-Y60591.  
 XX  
 SQ Sequence 215 BP; 66 A; 25 C; 58 G; 66 T; 0 other;

Query Match 1.7%; Score 18; DB 20; Length 215;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 736 aaatctctctctgagtc 753  
 |||||  
 Db 27 aaatctctctgagtc 44

RESULT 4  
 AAT79127  
 ID AAT79127 standard; cDNA to mRNA; 699 BP.  
 XX  
 AC AAT79127;  
 XX  
 DT 08-OCT-1997 (first entry)  
 XX  
 DE Human serine protease 60 (SP60) cDNA.  
 DE  
 KW Human; colon carcinoma; COLO 201; cell line; serine protease; SP60;  
 KW screening; inhibitor; treatment; disease; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT mat\_peptide 1..699 /\*tag= a  
 FT /product= serine\_protease  
 XX  
 FT JP09149790-A.  
 XX  
 PN 10-JUN-1997.  
 XX  
 XX 24-JUL-1996; 96JP-0212196.  
 XX  
 XX 29-SEP-1995; 95JP-0275105.  
 XX  
 XX (SUNR ) SUNTORY LTD.  
 XX  
 XX WPI; 1997-357902/33.  
 DR P-PSDB; AAW22986.  
 XX  
 XX Human colon carcinoma derived serine protease(s) SP59, SP60 and SP67  
 PT - useful to screen for specific inhibitors, e.g. to search for, or  
 PT study agent for treatment of various diseases  
 XX  
 PS Claim 2; Pages 11-12; 16pp; Japanese.  
 XX

CC The present sequence encodes the human colon carcinoma COLO 201  
 CC cell line derived serine protease 60 (SP60), which can be used to  
 CC screen for specific inhibitors, e.g. to search for, or study an

CC agent for the treatment of various diseases.  
 XX  
 SQ Sequence 699 BP; 157 A; 190 C; 209 G; 143 T; 0 other;

Query Match 1.7%; Score 18; DB 18; Length 699;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 tggccttgaggagtcagc 132  
 |||||  
 Db 34 tggccttgaggagtcagc 51

RESULT 5  
 AA242170  
 ID AA242170 standard; cDNA; 750 BP.  
 XX  
 AC AA242170;  
 XX  
 DT 31-JAN-2000 (first entry)  
 XX  
 DE Human normal bladder tissue cDNA derived EST 49.  
 DE  
 KW Human; bladder; treatment; EST; expressed sequence tag; cytostatic;  
 KW cancer; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN DE19818620-A1.  
 XX  
 XX 28-OCT-1999.  
 XX  
 XX 21-APR-1998; 98DE-1018620.  
 XX  
 XX 21-APR-1998; 98DE-1018620.  
 XX  
 XX (META-) METAGEN GES GENOMFORSCHUNG MBH.  
 XX  
 XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;  
 XX WPI; 1999-602416/52.  
 DR P-PSDB; AAY60398, AAY60399, AAY60400.  
 XX  
 XX New polypeptides and their nucleic acids, useful for treatment of  
 PT bladder tumour and identification of therapeutic agents -  
 XX  
 PS Claim 3; Page 184; 366pp; German.

XX This invention describes novel polypeptide fragment sequences (I) and  
 CC their encoding nucleic acids (II) which are highly expressed in normal  
 CC bladder tissue and have cytostatic activity. (II) are used for  
 CC recombinant expression of (I) and to isolate complete genes. (I) are  
 CC used to identify agents suitable for the treatment of bladder tumours,  
 CC to directly treat this form of cancer (including expression from gene  
 CC therapy vectors), or are used in a preparation for cancer treatment. (I)  
 CC is also used for the generation of specific antibodies. (II) are  
 CC identified by assembling ESTs (expressed sequence tags) from a  
 CC particular tissue type before comparison of expression patterns. This  
 CC allows a significantly longer fragment of the gene to be revealed, and  
 CC therefore reduces the number of failures because of ESTs from different  
 CC libraries representing different parts of the same unknown gene  
 CC distorting the estimated frequency of occurrence in a particular tissue.  
 CC AA242122-242248 represent EST fragments derived from a human normal  
 CC bladder tissue cDNA library which encode the protein fragments  
 CC represented in AAY60329-Y60591.  
 XX

SQ Sequence 750 BP; 193 A; 140 C; 157 G; 260 T; 0 other;

Query Match 1.7%; Score 18; DB 20; Length 750;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 736 aaatctctctcggagtc 753  
|||||

Db 197 aaatctctctcggagtc 214

RESULT 6

AAAF1559  
ID AAA61559 standard; cDNA; 1034 BP.

XX AC AAA61559;

XX XX 23-OCT-2000 (first entry)

XX DE cDNA encoding mouse trypsin family serine protease Tespec PRO-2.  
XX KW Tespec PRO-2; testis specific serine protease; mouse; murine;  
KW trypsin family serine protease; mature testis; sperm differentiation;  
KW sperm maturation; male infertility; sterility; reproductive disorder;  
KW contraception; ss.  
XX OS Mus musculus.

XX FH Key Location/Qualifiers  
XX FT CDS 69..1028  
XX FT /\*tag: a  
XX FT /product= "Mouse Tespec PRO-2"

XX PN WO200026352-A1.

XX PD 11-MAY-2000.

XX PF 02-NOV-1999; 99WO-JP06111.

XX PR 04-NOV-1998; 98JP-0313366.

XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX PI Senoo C, Numata M;

XX DR WPI; 2000-365604/31.

XX DR P-PSDB; AAB03157.

XX PT Trypsin family serine proteases expressed specifically in mature testis  
XX PT for development of methods for diagnosis and treatment of sterility and  
XX PT for contraception

XX PS Claim 2; Fig 2; 121pp; Japanese.

XX CC This sequence represents cDNA encoding mouse Tespec PRO-2 (testis  
XX CC specific serine protease-2). The invention relates to novel murine and  
XX CC human testis specific serine proteases (AAB03156-R03160) and to cDNAs  
XX CC encoding them (AAA61558-A61562). It also encompasses expression vectors  
XX CC and host cells comprising a nucleotide sequence encoding a protease of  
XX CC the invention, inhibitors of the proteases and antibodies against the  
XX CC proteases. The novel proteases are members of the trypsin family of  
XX CC serine proteases, having the serine and histidine active site signatures  
XX CC characteristic of this family. The proteases are specifically expressed  
XX CC in mature testis and participate in the differentiation and maturation  
XX CC of sperm. The proteases are potentially useful for the development of  
XX CC pharmaceuticals for the treatment of male infertility and other male  
XX CC reproductive disorders, and for the development of contraceptives. They  
XX CC may also be used as reagents for the diagnosis of male infertility.

XX SQ Sequence 1034 BP; 262 A; 267 C; 277 G; 227 T; 1 other;

Query Match 1.7%; Score 18; DB 21; Length 1034;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 ctgacagcagcactgc 207

|||||

Db 294 ctgacagcagcactgc 311

RESULT 7

AAF24159  
ID AAF24159 standard; DNA; 1232 BP.

XX AC AAF24159;

XX XX 23-MAR-2001 (first entry)

XX DE Human secreted protein DNA #9.

XX KW Secreted protein; gene therapy; vaccine; cancer; leukemia;  
KW autoimmune disease; allergy; inflammation; graft rejection;  
KW hyperproliferation; cardiovascular; infection; ds.

XX OS Homo sapiens.

XX PN WO200075375-A1.

XX PD 14-DEC-2000.

XX PF 02-JUN-2000; 2000WO-US15187.

XX PR 07-JUN-1999; 99US-0137725.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Birse CE, Duan RD, Soppet DR, Rosen CA, Shi Y;  
PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Ni J, Young PE;  
XX WPI; 2001-061741/07.

XX CC Nucleic acids encoding 26 human secreted polypeptides, useful e.g. for  
XX CC preventing, diagnosing and/or treating cancers and for promoting wound  
XX CC healing

XX PS Claim 1; Page 423-424; 530pp; English.

XX CC The present invention relates to 26 secreted human proteins. The  
XX CC proteins may be used in the prevention, diagnosis and treatment of  
XX CC diseases associated with inappropriate polypeptide expression.  
XX CC For example, they may be used in gene therapy or in vaccines.  
XX CC Typical of diseases which are potentially treatable are cancers  
XX CC (including leukemia), autoimmune diseases, allergies, inflammation,  
XX CC graft rejection, hyperproliferation, cardiovascular diseases  
XX CC (particularly critical limb ischemia and coronary disease) and any  
XX CC involving abnormal angiogenesis, neurodegeneration and/or  
XX CC infectious diseases.

XX SQ Sequence 1232 BP; 333 A; 249 C; 306 G; 342 T; 2 other;

Query Match 1.7%; Score 18; DB 22; Length 1232;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 cgctgctctcttctgctgg 43

|||||

Db 74 cgctgctctcttctgctgg 91

RESULT 8

AAF24181  
ID AAF24181 standard; DNA; 1368 BP.

XX AC AAF24181;

XX XX 23-MAR-2001 (first entry)

XX DE Human secreted protein DNA #19.

XX XX



KW Secreted protein; gene therapy; vaccine; cancer; leukemia;  
 KW autoimmune disease; allergy; inflammation; graft rejection;  
 KW hyperproliferation; cardiovascular; infection; ds.  
 OS Homo sapiens.  
 XX WO200075375-A1.  
 PN 14-DEC-2000.  
 XX 02-JUN-2000; 2000WO-US15187.  
 PF 07-JUN-1999; 99US-0137725.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 XX Ruben SM, Birse CE, Duan RD, Soppet DR, Rosen CA, Shi Y;  
 PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Ni J, Young PE;  
 XX WPI; 2001-061741/07.  
 DR Nucleic acids encoding 26 human secreted polypeptides, useful e.g. for  
 XX preventing, diagnosing and/or treating cancers and for promoting wound  
 XX healing -  
 XX Claim 1; Page 441-442; 530pp; English.  
 XX The present invention relates to 26 secreted human proteins. The  
 CC proteins may be used in the prevention, diagnosis and treatment of  
 CC diseases associated with inappropriate polypeptide expression.  
 CC For example, they may be used in gene therapy or in vaccines.  
 CC Typical of diseases which are potentially treatable are cancers  
 CC (including leukemia), autoimmune diseases, allergies, inflammation,  
 CC graft rejection, hyperproliferation, cardiovascular diseases  
 CC (particularly critical limb ischemia and coronary disease) and any  
 CC involving abnormal angiogenesis, neurodegeneration and/or  
 CC infectious diseases.  
 XX Sequence 1268 BP; 355 A; 271 C; 316 G; 326 T; 0 other;

Query Match 1.7%; Score 18; DB 22; Length 1268;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 cgctgctctcttgcctgg 43  
 Db 68 cgctgctctcttgcctgg 85

RESULT 9  
 AAF24182  
 ID AAF24182 standard; DNA; 1268 BP.

XX AAF24182;  
 XX 23-MAR-2001 (first entry)  
 XX Human secreted protein DNA #19.

XX Secreted protein; gene therapy; vaccine; cancer; leukemia;  
 KW autoimmune disease; allergy; inflammation; graft rejection;  
 KW hyperproliferation; cardiovascular; infection; ds.

OS Homo sapiens.  
 XX WO200075375-A1.  
 PN 14-DEC-2000.  
 XX 02-JUN-2000; 2000WO-US15187.  
 PF 07-JUN-1999; 99US-0137725.

XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Ruben SM, Birse CE, Duan RD, Soppet DR, Rosen CA, Shi Y;  
 PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Ni J, Young PE;  
 XX WPI; 2001-061741/07.  
 DR Nucleic acids encoding 26 human secreted polypeptides, useful e.g. for  
 XX preventing, diagnosing and/or treating cancers and for promoting wound  
 XX healing -  
 XX Claim 1; Page 443; 530pp; English.  
 XX The present invention relates to 26 secreted human proteins. The  
 CC proteins may be used in the prevention, diagnosis and treatment of  
 CC diseases associated with inappropriate polypeptide expression.  
 CC For example, they may be used in gene therapy or in vaccines.  
 CC Typical of diseases which are potentially treatable are cancers  
 CC (including leukemia), autoimmune diseases, allergies, inflammation,  
 CC graft rejection, hyperproliferation, cardiovascular diseases  
 CC (particularly critical limb ischemia and coronary disease) and any  
 CC involving abnormal angiogenesis, neurodegeneration and/or  
 CC infectious diseases.  
 XX Sequence 1268 BP; 354 A; 271 C; 316 G; 327 T; 0 other;

Query Match 1.7%; Score 18; DB 22; Length 1268;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 cgctgctctcttgcctgg 43  
 Db 68 cgctgctctcttgcctgg 85

RESULT 10  
 AAD02556  
 ID AAD02556 standard; cDNA; 1305 BP.

XX AAD02556;  
 XX 02-MAY-2001 (first entry)  
 XX Human seripancrin cDNA.

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;  
 KW arthritis; chronic obstructive pulmonary disorder; COPD; cancer;  
 KW osteoporosis; aberrant wound healing; angiogenesis; diabetes;  
 KW inflammatory disorder; stroke; cardiovascular disease; gene therapy;  
 KW vaccine; cytostatic; cerebroprotective; vulnery; osteopathic; ss.

OS Homo sapiens.

XX Key Location/Qualifiers  
 XX CDS 1..1305  
 XX FT /\*tag= a  
 XX FT /product= "Human seripancrin protein"  
 XX FT /note= "The coding region does not include stop codon"  
 XX FT /partial

XX WO200104141-A2.  
 XX 18-JAN-2001.  
 XX 04-JUL-2000; 2000WO-EP06211.  
 XX 12-JUL-1999; 99EP-0113428.  
 XX (MERE ) MERCK PATENT GMBH.  
 XX Suendermann B, Hofmann U, Matzku S, Wilbert O;

```

XX WPI: 2001-147177/15.
DR P-PSDB; AAY72558.
XX
PT New extracellular serine protease Seripancrin, useful for treating
PT cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease,
PT diabetes, inflammatory disorders, stroke, angiogenesis and aberrant
PT wound healing.
XX
XX Claim 5; Page 37-39; 45pp; English.
XX
CC The present invention relates to seripancrin polynucleotides, and
CC polypeptides encoded by them. Seripancrin are members of serine
CC protease family. This protein contains a transmembrane domain,
CC a low density lipoprotein (LDL) domain, a protease domain and a
CC scavenger receptor cysteine-rich (SRCR) domain. The LDL and SRCR
CC domains help to define the specificity of seripancrin's intra and
CC intermolecular interactions. The polynucleotides and polypeptides
CC of the invention are useful for treating and diagnosing diseases
CC such as arthritis, chronic obstructive pulmonary disorder (COPD),
CC cancer, osteoporosis, aberrant wound healing, angiogenesis,
CC inflammatory disorders, diabetes, stroke and cardiovascular diseases.
CC Seripancrin genes are useful in chromosome localisation studies,
CC as tools for tissue expression studies and also in gene therapy.
CC The polypeptides of the invention are used for identifying agonists
CC and antagonists useful for treating conditions associated with
CC seripancrin imbalance. These polypeptides are also useful as vaccines.
CC The present sequence is a cDNA coding for seripancrin protein. The
CC seripancrin gene is located on human chromosome 11q22-q23.
XX
XX Sequence 1305 BP; 293 A; 368 C; 372 G; 272 T; 0 other;
XX
XX
XX Query Match 1.7%; Score 18; DB 22; Length 1305;
XX Best Local Similarity 100.0%; Pred. No. 28;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 115 tggccttgccaggtcagc 132
Db 640 tggccttgccaggtcagc 657
XXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXX
RESULT 11
AAD01122
ID AAD01122 standard; cDNA; 1413 BP.
XX
XX AAD01122;
XX
XX 02-NOV-2000 (first entry)
XX
XX Human orphan G protein-coupled receptor hrUP5 cDNA.
XX
XX Human; orphan G protein-coupled receptor; GPCR; hrUP5; drug screening;
XX transmembrane receptor; signal cascade; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1413
XX /*tag= a
XX /product= "hrUP5"
XX /note= "Human orphan G protein-coupled receptor"
XX /transl_except= (pos:652..660, aa:Thr)
XX
XX WO200031258-A2.
XX
XX 02-JUN-2000.
XX
XX 13-OCT-1999; 99WO-US23687.
XX
XX 20-NOV-1998; 98US-0109213.
XX 16-FEB-1999; 99US-0120416.
XX 26-FEB-1999; 99US-0121852.
XX

```

```

PR 12-MAR-1999; 99US-0123946.
PR 12-MAR-1999; 99US-0123949.
PR 28-MAY-1999; 99US-0136436.
PR 28-MAY-1999; 99US-0136437.
PR 28-MAY-1999; 99US-0136439.
PR 28-MAY-1999; 99US-0136567.
PR 28-MAY-1999; 99US-0137127.
PR 28-MAY-1999; 99US-0137131.
PR 29-JUN-1999; 99US-0141448.
PR 29-SEP-1999; 99US-0156555.
PR 29-SEP-1999; 99US-0156633.
PR 29-SEP-1999; 99US-0156634.
PR 29-SEP-1999; 99US-0156653.
PR 01-OCT-1999; 99US-0157280.
PR 01-OCT-1999; 99US-0157281.
PR 01-OCT-1999; 99US-0157282.
PR 01-OCT-1999; 99US-0157293.
PR 01-OCT-1999; 99US-0157294.
PR 12-OCT-1999; 99US-0416760.
PR 12-OCT-1999; 99US-0417044.
XX
XX (AREN-) ARENA PHARM INC.
XX
XX Chen R, Dang HT, Liaw CW, Lin I;
XX WPI; 2000-400068/34.
XX P-PSDB; AAY71295.
XX
XX Novel human orphan G protein-coupled receptors and the encoding cDNAs
XX for use in the identification of G protein-coupled receptor agonists -
XX Claim 17; Page 53-54; 102pp; English.
XX
XX The present sequence is a cDNA encoding hrUP5, an endogenous human
XX orphan G protein-coupled receptor (GPCR). The full length hrUP5 cDNA was
XX cloned by RT-PCR using human peripheral leucocyte cDNA as template.
XX The orphan GPCR of the invention, like all GPCRs has seven transmembrane
XX alpha helices with an extracellular N-terminus and an intracellular
XX C-terminus. However, no endogenous ligands has yet been identified for
XX the proteins of the invention. The orphan GPCRs may be used in the
XX identification of their endogenous ligands, and to screen potential GPCR
XX agonists and antagonists for use as pharmaceutical agents. The proteins
XX may also be used in the study of GPCR-mediated signalling cascades, and
XX to elucidate their precise role in normal and diseased human conditions.
XX Nucleic acid encoding human orphan GPCRs may be used for tissue
XX localisation expression analysis to provide information about their
XX function in healthy and pathological states.
XX
XX Sequence 1413 BP; 212 A; 537 C; 393 G; 271 T; 0 other;
XX
XX
XX Query Match 1.7%; Score 18; DB 21; Length 1413;
XX Best Local Similarity 100.0%; Pred. No. 28;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 821 tctctgactcttctgttcc 838
Db 224 tctctgactcttctgttcc 241
XXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXX
RESULT 12
AAD46021
ID AAD46021 standard; cDNA; 1413 BP.
XX
XX AAD46021;
XX
XX 22-AUG-2000 (first entry)
XX
XX Human G protein coupled receptor hrUP5 encoding cDNA SEQ ID NO:9.
XX
XX Human; G protein coupled receptor; GPCR; transmembrane receptor;
XX identification; agonist; screening; therapeutic; pharmaceutical;
XX mutant; ss.
XX

```

```
XX OS Homo sapiens.
XX PN WO200022131-A2.
XX PD 20-APR-2000.
XX PF 13-OCT-1999; 99WO-US24065.
XX PR 13-OCT-1998; 98US-0170496.
XX PR 12-NOV-1998; 98US-0108029.
XX PR 20-NOV-1998; 98US-0109213.
XX PR 27-NOV-1998; 98US-0110060.
XX PR 16-FEB-1999; 99US-0120416.
XX PR 26-FEB-1999; 99US-0121852.
XX PR 12-MAR-1999; 99US-0123944.
XX PR 12-MAR-1999; 99US-0123945.
XX PR 12-MAR-1999; 99US-0123946.
XX PR 12-MAR-1999; 99US-0123948.
XX PR 12-MAR-1999; 99US-0123949.
XX PR 12-MAR-1999; 99US-0123951.
XX PR 28-MAY-1999; 99US-0136436.
XX PR 28-MAY-1999; 99US-0136437.
XX PR 28-MAY-1999; 99US-0136439.
XX PR 28-MAY-1999; 99US-0137127.
XX PR 28-MAY-1999; 99US-0137131.
XX PR 28-MAY-1999; 99US-0137567.
XX PR 30-JUN-1999; 99US-0141448.
XX PR 27-AUG-1999; 99US-0151114.
XX PR 03-SEP-1999; 99US-0152524.
XX PR 29-SEP-1999; 99US-0156633.
XX PR 29-SEP-1999; 99US-0156555.
XX PR 29-SEP-1999; 99US-0156634.
XX PA (AREN-) ARENA PHARM INC.
XX PI Behan DP, Lehmann-Bruinsma K, Chalmers DT, Chen R, Dang HT;
XX PI Gore M, Liaw CW, Lin I, Lowitz K, White C;
XX DR WPI: 2000-317986/27.
XX DR P-FSDB: AAB02829.
XX PT Non-endogenous, human G protein-coupled receptors for screening
XX PT receptor, inverse or partial agonists useful as therapeutic agents -
XX PS Example 1; Page 82-83; 187pp; English.
XX CC The present invention describes transmembrane receptors, preferably
XX CC human G protein coupled receptors (GPCR), for which the endogenous
XX CC ligand is unknown (orphan GPCR receptors). More specifically the present
XX CC invention relates to non-endogenous, constitutively activated versions
XX CC of a human GPCR. These non-endogenous human GPCRs can be useful for
XX CC the direct identification of candidate compounds as receptors agonists,
XX CC inverse agonists or partial agonists for use as pharmaceutical agents.
XX CC AAA46017 to AAA46126 and AAB02825 to AAB02859 represent sequences used in
XX CC the exemplification of the present invention.
XX SQ Sequence 1413 BP; 212 A; 537 C; 393 G; 271 T; 0 other;

Query Match 1.7%; Score 18; DB 21; Length 1413;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 821 tctctgactcttctgtcc 838
Db 224 tctctgactcttctgtcc 241

RESULT 13
AAAI5587
ID AAAI5587 standard; cDNA; 1413 BP.
XX
```

```
AC AAAI5587;
XX 01-AUG-2000 (first entry)
XX DE Human G-protein coupled receptor, AXOR14, coding sequence.
XX KW Human; G-protein coupled receptor; AXOR14; signal transduction;
XX KW 7TM receptor; gene therapy; infection; cancer; autoimmunity;
XX KW Parkinson's disease; cardiovascular disorder; neurological disorder;
XX KW Huntington's disease; diabetes; obesity; dyskinesias; chromosome 11q13;
XX KW anorexia; bulimia; osteoporosis; 7 transmembrane receptor; EST;
XX KW expressed sequence tag; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 1..1413
XX FT /*tag= a
XX FT /product= "AXOR14"
XX PN WO200026339-A2.
XX PD 11-MAY-2000.
XX PF 02-NOV-1999; 99WO-US25791.
XX PR 03-NOV-1998; 98GB-0024027.
XX PR 02-MAR-1999; 99US-0260298.
XX PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX PI Elshourbagy N, Michalovich D;
XX DR WPI: 2000-365593/31.
XX DR P-FSDB: AAY94268.
XX PT New AXOR14 polypeptides and polynucleotides useful for treating e.g.
XX PT microbial infections, pain, cancers, psychotoc and neurological
XX PT disorders, allergies
XX PS Claim 13; Page 37; 38pp; English.
XX CC The present sequence is the human G-protein coupled receptor, AXOR14,
XX CC gene as derived from an Expressed Sequence Tag (EST). G-protein coupled
XX CC receptors are also known as 7-transmembrane (7TM) receptors. The present
XX CC sequence is located on chromosome 11q13. The AXOR14 protein functions in
XX CC hormone signal transduction. AXOR14 protein may be used in the
XX CC identification of agonists, antagonists or inhibitors that can be
XX CC used in therapy of microbial infections (e.g. HIV-1 and HIV-2), pain,
XX CC cancers, psychotoc and neurological disorders, allergies, diabetes,
XX CC obesity, anorexia, bulimia, Parkinson's disease, acute heart failure,
XX CC hypotension, hypertension, urinary retention, osteoporosis, angina
XX CC pectoris, myocardial infarction, stroke, benign prostatic hypertrophy,
XX CC vomiting, dyskinesias or Huntington's disease which may be caused by
XX CC inappropriate AXOR14 activity or imbalance. The actual gene may also be
XX CC used in gene therapy for the above disorders.
XX SQ Sequence 1413 BP; 212 A; 537 C; 393 G; 271 T; 0 other;

Query Match 1.7%; Score 18; DB 21; Length 1413;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 821 tctctgactcttctgtcc 838
Db 224 tctctgactcttctgtcc 241

RESULT 14
AAAI5587
ID AAD02557 standard; DNA; 1479 BP.
XX
```

```
AC AAD02557;
XX
XX 02-MAY-2001 (first entry)
XX
XX Human seripancrin variant #1 DNA.
XX
XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;
KW arthritis; Chronic obstructive pulmonary disorder; COPD; cancer;
KW osteoporosis; aberrant wound healing; angiogenesis; diabetes;
KW inflammatory disorder; stroke; cardiovascular disease; gene therapy;
KW vaccine; cytostatic; cerebroprotective; vulnerary; osteopathic; ds.
XX
XX Homo sapiens.
XX
XX
XX Key Location/Qualifiers
FH CDS 1..1479
FT /*tag= a
FT /product= "Human seripancrin variant #1 protein"
XX
XX WO200104141-A2.
XX
XX 18-JAN-2001.
XX
XX 04-JUL-2000; 2000WO-EP06211.
XX
XX 12-JUL-1999; 99EP-0113428.
XX
XX (MERE ) MERCK PATENT GMBH.
XX
XX Suendermann B, Hofmann U, Matzku S, Wilbert O;
XX
XX WPI: 2001-147177/15.
XX P-PSDB; AAY72559.
XX
XX New extracellular serine protease Seripancrin, useful for treating
PT cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease,
PT diabetes, inflammatory disorders, stroke, angiogenesis and aberrant
PT wound healing
XX
XX Disclosure; Page 40-42; 45pp; English.
XX
XX The present invention relates to seripancrin polynucleotides, and
CC polypeptides encoded by them. Seripancrin are members of serine
CC protease family. This protein contains a transmembrane domain,
CC a low density lipoprotein (LDL) domain, protease domain and a
CC scavenger receptor cysteine-rich (SRCR) domain. The LDL and SRCR
CC domains help to define the specificity of seripancrin's intra and
CC intermolecular interactions. The polynucleotides and polypeptides
CC of the invention are useful for treating and diagnosing diseases
CC such as arthritis, chronic obstructive pulmonary disorder (COPD),
CC cancer, osteoporosis, aberrant wound healing, angiogenesis,
CC inflammatory disorders, diabetes, stroke and cardiovascular diseases.
CC Seripancrin genes are useful in chromosome localisation studies,
CC as tools for tissue expression studies and also in gene therapy.
CC The polypeptides of the invention are used for identifying agonists
CC and antagonists useful for treating conditions associated with
CC seripancrin imbalance. These polypeptides are also useful as vaccines.
CC The present sequence is a DNA coding for seripancrin variant #1
CC protein. The seripancrin gene is located on human chromosome 11q22-q23.
XX
XX Sequence 1479 BP; 329 A; 425 C; 407 G; 318 T; 0 other;
SQ
Query Match 1.7%; Score 18; DB 22; Length 1479;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 115 tggccttgccaggtcagc 132
DB 640 tggccttgccaggtcagc 657
RESULT * 15
```

```
AAA07582/C
ID AAA07582 standard; DNA; 1959 BP.
XX
XX AAA07582;
XX
XX 29-AUG-2000 (first entry)
XX
XX Marigold beta-cyclase coding sequence.
XX
XX Marigold; beta-cyclase; beta-hydroxylase; epsilon-cyclase; enzyme;
KW IPP isomerase; isopentyl pyrophosphate; transgenic plant;
KW catalytic activity; carotenoid production; ds.
XX
XX Tagetes erecta.
XX
XX Key Location/Qualifiers
FH CDS 304..1839
FT /*tag= a
FT /product= beta-cyclase
XX
XX WO200032788-A2.
XX
XX 08-JUN-2000.
XX
XX 30-NOV-1999; 99WO-DK00668.
XX
XX 30-NOV-1998; 98US-0201641.
XX
XX (CHRH-) CHR HANSEN AS.
XX
XX Dellapenna D, Cunningham FX;
XX
XX WPI: 2000-412334/35.
XX P-PSDB; AAY90226.
XX
XX Novel method for regulating carotenoid biosynthesis in Marigolds by
PT modulating the catalytic activity of beta- cyclase, beta-hydroxylase,
PT epsilon cyclase, or IPP isomerase
XX
XX Claim 18; Fig 3; 69pp; English.
XX
XX This sequence encodes the marigold beta-cyclase. The invention relates to
CC transgenic plant material containing an isolated DNA encoding a marigold
CC enzyme having catalytic activity of beta-cyclase, beta-hydroxylase,
CC epsilon-cyclase, or isopentyl pyrophosphate (IPP) isomerase. The DNA
CC sequences can be used in a method for manipulating carotenoid synthesis
CC in a plant material. The method is used to produce transgenic marigold
CC plants which have enhanced production of specific carotenoid products,
CC e.g. beta-carotene, alpha-carotene, zeaxanthin, lycopene, zeinoxanthin,
CC beta-cryptoxanthin, rubixanthin, and combinations thereof. The method
CC can also be used to increase or decrease expression of beta-cyclase,
CC beta-hydroxylase, epsilon cyclase, or IPP isomerase enzymes in transgenic
CC marigolds. The invention allows enhanced production of specific
CC carotenoid products.
XX
XX Sequence 1959 BP; 583 A; 312 C; 418 G; 646 T; 0 other;
SQ
Query Match 1.7%; Score 18; DB 21; Length 1959;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 997 gagccactgctaaccctg 1014
DB 622 GAGCCACTGCTAACCCCTG 605
RESULT 16
AAX87154
ID AAX87154 standard; cDNA; 2038 BP.
XX
XX AAX87154;
XX
```

DT 27-SEP-1999 (first entry)  
XX Human protease HUPM-6 cDNA.  
DE  
XX Serine protease; human; HUPM-6; cell proliferation; cancer;  
KW immune disorder; inflammation; therapy; ss.  
KW  
XX Homo sapiens.  
OS  
XX Key Location/Qualifiers  
FH 200...1507  
FT CDS /\*tag= a  
FT  
XX WO9936550-A2.  
PN  
XX 22-JUL-1999.  
PD  
XX 12-JAN-1999; 99WO-US00655.  
PF  
XX 16-JAN-1998; 98US-0008271.  
PR  
XX (INCY-) INCYTE PHARM INC.  
PA  
XX Bandman O, Corley NC, Guegler KJ, Hillman JL, Shah P;  
PI Tang YT, Yue H;  
PI  
XX WPI: 1999-430616/36.  
DR  
XX P-PSDB; AAY06437.  
DR  
XX Novel human protease molecules useful in the treatment of  
PT developmental disorders and/or cancers  
PT  
XX Claim 8; Page 86-87; 90pp; English.  
PS  
XX This nucleotide sequence codes for HUPM-6 (see AAY06437), a novel  
CC human protease. HUPM-6 cDNA was initially identified in Incyte  
CC Clone 1337018 from the colon cDNA library COLNNT013 using a  
CC computer search for amino acid sequence alignments. The present  
CC sequence is a consensus sequence derived from overlapping and/or  
CC extended nucleic acid sequences: Incyte clones 1271725 (TESTTUT02),  
CC 1337018, 586982 and 588598 (UTRSNOT01). A fragment comprising  
CC nucleotides 900-949 of the present sequence can be used for  
CC hybridisation. This sequence encompasses an active site residue.  
CC Northern analysis shows expression of HUPM-6 in gastrointestinal,  
CC and male and female reproductive cDNA libraries. Approximately 65%  
CC of these libraries are associated with neoplastic disorders, and  
CC 22% with the immune response. The invention provides 12 new human  
CC proteases, i.e. HUPM-1 to -12 (see AAY06432-43), and the  
CC polynucleotides encoding them (see AAX87149-60). Also provided are  
CC vectors, host cells and methods for producing HUPM polypeptides, as  
CC well as agonists and antagonists of HUPM. Methods for treating or  
CC preventing cell proliferative disorders and immune disorders using  
CC HUPM or HUPM antagonists are claimed.  
XX\*  
SQ Sequence 2038 BP; 462 A; 591 C; 569 G; 416 T; 0 other;  
  
Query Match 1.7%; Score 18; DB 20; Length 2038;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 115 tgcccttgccagcagc 132  
|||||  
Db 839 tgcccttgccagcagc 856  
  
RESULT 17  
AAA37099  
ID AAA37099 standard; cDNA; 2063 BP.  
XX  
AC AAA37099;  
XX  
DT 08-AUG-2000 (first entry)

XX Human PRO1570 (UNC776) cDNA sequence SEQ ID NO:274.  
DE  
XX Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;  
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening;  
KW ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200012708-A2.  
PN  
XX 09-MAR-2000.  
PD  
XX  
XX 01-SEP-1999; 99WO-US20111.  
PF  
XX 01-SEP-1998; 98US-0098716.  
PR 01-SEP-1998; 98US-0098749.  
PR 01-SEP-1998; 98US-0098750.  
PR 02-SEP-1998; 98US-0098803.  
PR 02-SEP-1998; 98US-0098821.  
PR 02-SEP-1998; 98US-0098843.  
PR 02-SEP-1998; 98US-0099536.  
PR 02-SEP-1998; 98US-0099596.  
PR 09-SEP-1998; 98US-0099598.  
PR 09-SEP-1998; 98US-0099602.  
PR 09-SEP-1998; 98US-0099642.  
PR 10-SEP-1998; 98US-0099741.  
PR 10-SEP-1998; 98US-0099754.  
PR 10-SEP-1998; 98US-0099763.  
PR 10-SEP-1998; 98US-0099792.  
PR 10-SEP-1998; 98US-0099808.  
PR 10-SEP-1998; 98US-0099812.  
PR 10-SEP-1998; 98US-0099815.  
PR 10-SEP-1998; 98US-0099816.  
PR 13-SEP-1998; 98US-0100385.  
PR 15-SEP-1998; 98US-0100388.  
PR 15-SEP-1998; 98US-0100390.  
PR 16-SEP-1998; 98US-0100584.  
PR 16-SEP-1998; 98US-0100627.  
PR 16-SEP-1998; 98US-0100661.  
PR 16-SEP-1998; 98US-0100662.  
PR 16-SEP-1998; 98US-0100664.  
PR 17-SEP-1998; 98US-0100683.  
PR 17-SEP-1998; 98US-0100684.  
PR 17-SEP-1998; 98US-0100710.  
PR 17-SEP-1998; 98US-0100711.  
PR 17-SEP-1998; 98US-0100919.  
PR 17-SEP-1998; 98US-0100930.  
PR 18-SEP-1998; 98US-0100848.  
PR 18-SEP-1998; 98US-0100849.  
PR 18-SEP-1998; 98US-0101014.  
PR 18-SEP-1998; 98US-0101068.  
PR 18-SEP-1998; 98US-0101071.  
PR 18-SEP-1998; 98US-0101071.  
PR 22-SEP-1998; 98US-0101279.  
PR 23-SEP-1998; 98US-0101471.  
PR 23-SEP-1998; 98US-0101472.  
PR 23-SEP-1998; 98US-0101474.  
PR 23-SEP-1998; 98US-0101475.  
PR 23-SEP-1998; 98US-0101476.  
PR 23-SEP-1998; 98US-0101477.  
PR 23-SEP-1998; 98US-0101479.  
PR 24-SEP-1998; 98US-0101738.  
PR 24-SEP-1998; 98US-0101741.  
PR 24-SEP-1998; 98US-0101743.  
PR 24-SEP-1998; 98US-0101915.  
PR 24-SEP-1998; 98US-0101916.  
PR 29-SEP-1998; 98US-0102207.  
PR 29-SEP-1998; 98US-0102240.  
PR 29-SEP-1998; 98US-0102307.  
PR 29-SEP-1998; 98US-0102330.  
PR 29-SEP-1998; 98US-0102331.  
PR 30-SEP-1998; 98US-0102484.  
PR 30-SEP-1998; 98US-0102487.



PS Claim 2; Fig 111; 278pp; English.

XX The present sequence is the coding sequence for a human PRO polypeptide

CC (secreted and transmembrane). The PRO protein, and PRO agonists, PRO

CC antagonists or anti-PRO antibodies are useful for preparation of a

CC medicament useful in the treatment of a condition which is responsive to

CC the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO

CC protein may also be employed as molecular weight markers for protein

CC electrophoresis. The PRO coding sequence has applications in molecular

CC biology, including use as hybridisation probes, and in chromosome and

XX gene mapping.

SQ Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

Query Match 1.7%; Score 18; DB 22; Length 2063;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 tggccttgccaggtcagc 132

Db 840 tggccttgccaggtcagc 857

RESULT 19

AAF54396

ID AAF54396 standard; DNA; 2063 BP.

XX AAF54396;

XX 02-APR-2001 (first entry)

DT DNA encoding protein of the invention #76.

DE Secreted; transmembrane; gene therapy; ss.

XX Unidentified.

OS WO200078961-A1.

PN 28-DEC-2000.

PD 18-FEB-2000; 2000WO-US04342.

XX 23-JUN-1999; 99US-0141037.

PR 20-JUL-1999; 99US-0144758.

PR 26-JUL-1999; 99US-0145698.

PR 01-SEP-1999; 99WO-US20111.

PR 29-OCT-1999; 99US-0162506.

PR 30-NOV-1999; 99WO-US28313.

PR 02-DEC-1999; 99WO-US28551.

PR 16-DEC-1999; 99WO-US30095.

PR 05-JAN-2000; 2000WO-US00219.

PR 06-JAN-2000; 2000WO-US00376.

XX (GETH ) GENENTECH INC.

PA Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;

PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

PI Fan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

PI Watanabe CK, Williams PM, Wood WI;

XX WPI; 2001-071395/08.

DR Secreted and transmembrane proteins and nucleic acids designated PRO,

XX useful as hybridization probes, in chromosome and gene mapping and gene

PT therapy -

XX Claim 2; Fig 151; 787pp; English.

PS The present invention relates to secreted and transmembrane proteins.

XX These proteins and the DNA encoding them may be used as hybridization

CC probes, in chromosome and gene mapping and in the generation of

CC

CC anti-sense RNA and DNA. They may also be used to generate either

CC transgenic animals or knockout animals which are in turn useful for

CC development and screening of therapeutically useful reagents.

CC The nucleic acids may also be used in gene therapy.

XX Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

SQ

Query Match 1.7%; Score 18; DB 22; Length 2063;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 tggccttgccaggtcagc 132

Db 840 tggccttgccaggtcagc 857

RESULT 20

AAZ90471

ID AAZ90471 standard; cDNA; 2070 BP.

XX AAZ90471;

XX 06-JUN-2000 (first entry)

DT Cancer specific gene (CSG) sequence (clone ID 1283171).

DE CSG; cancer specific gene; cancer; gynecologic cancer; ovarian; breast;

XX endometrial; uterine; lung; cytotoxic; ss.

OS Homo sapiens.

XX WO200012758-A1.

PN 09-MAR-2000.

PD 01-SEP-1999; 99WO-US19655.

PR 02-SEP-1998; 98US-0098880.

XX (DIAD-) DIADEXUS LLC.

PA Salceda S, Sun Y, Recipon H, Caferkey R;

PI WPI; 2000-256657/22.

DR Diagnosing, staging, monitoring, imaging and treating cancer especially

XX gynecological cancers e.g. breast, ovarian cancer and lung cancer,

PT involves measuring cancer specific gene levels in cells and body fluids

XX

PS Claim 9; Page 45-46; 58pp; English.

XX The invention relates to detecting, diagnosing metastasis and staging

CC cancer by measuring levels of cancer specific genes (CSG) in cells,

CC tissues or body fluids. Their remission and progression, decreases and

CC increases in CSG levels, is also monitored, by periodic sample analysis.

CC The methods are useful for detecting cancers, especially gynecologic

CC cancers which include ovarian, breast, endometrial and uterine cancer

CC and lung cancer. Antibodies against the CSGs labeled with paramagnetic

CC ions or a radioisotope is useful for imaging cancer and when conjugated

CC with a cytotoxic agent are useful for treating cancer. The present

CC sequence represents a CSG sequence (clone ID: 1283171 and

CC gene ID: 332459).

XX

SQ Sequence 2070 BP; 473 A; 599 C; 577 G; 421 T; 0 other;

Query Match 1.7%; Score 18; DB 21; Length 2070;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 tggccttgccaggtcagc 132

Db 859 tggccttgccaggtcagc 876  
|||||

RESULT 21  
AAAL5586  
ID AAA15586 standard; cDNA; 2100 BP.  
XX  
AC AAA15586;  
XX  
DT 01-AUG-2000 (first entry)  
XX  
DE Human G-protein coupled receptor, AXOR14, cDNA sequence.  
XX  
KW Human; G-protein coupled receptor; AXOR14; signal transduction;  
KW 7TM receptor; gene therapy; infection; cancer; autoimmunity;  
KW Parkinson's disease; cardiovascular disorder; neurological disorder;  
KW Huntington's disease; diabetes; obesity; dyskinesias; chromosome 11q13;  
KW anorexia; bulimia; osteoporosis; 7 transmembrane receptor; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 195..1607  
FT FT /\*tag= a  
FT FT /product= "AXOR14"  
XX  
PN WO200026339-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-US25791.  
XX  
PR 03-NOV-1998; 98GB-0024027.  
PR 02-MAR-1999; 99US-0260298.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
XX  
PI Elshourbagy N, Michalovich D;  
XX  
XX WPI; 2000-365593/31.  
DR P-PSDB; AAY94267.  
XX  
PT New AXOR14 polypeptides and polynucleotides useful for treating e.g.  
PT microbial infections, pain, cancers, psychotic and neurological  
PT disorders, allergies  
XX  
PS Claim 2; Page 35-36; 38pp; English.  
XX  
CC The present sequence is the human G-protein coupled receptor, AXOR14,  
CC gene. G-protein coupled receptors are also known as 7-transmembrane  
CC receptors. The present sequence is located on chromosome 11q13. The  
CC AXOR14 protein functions in hormone signal transduction. AXOR14 protein  
CC that can be used in the identification of agonists, antagonists or inhibitors  
CC that can be used in therapy of microbial infections (e.g. HIV-1 and  
CC HIV-2), pain, cancers, psychotic and neurological disorders, allergies,  
CC diabetes, obesity, anorexia, bulimia, Parkinson's disease, acute heart  
CC failure, hypotension, hypertension, urinary retention, osteoporosis,  
CC angina pectoris, myocardial infarction, stroke, benign prostatic  
CC hypertrophy, vomiting, dyskinesias or Huntington's disease which may be  
CC caused by inappropriate AXOR14 activity or imbalance. The actual gene  
CC may also be used in gene therapy for the above disorders.  
XX  
SQ Sequence 2100 BP; 376 A; 759 C; 573 G; 392 T; 0 other;

Query Match 1.7%; Score 18; DB 21; Length 2100;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 821 tctctgactcttctgttc 838  
Db 418 tctctgactcttctgttc 435

RESULT 22  
AAF26183  
ID AAF26183 standard; DNA; 3540 BP.  
XX  
AC AAF26183;  
XX  
DT 26-APR-2001 (first entry)  
XX  
DE Human G-protein coupled receptor N-formyl DNA.  
XX  
KW G-protein coupled receptor; N-formyl receptor; transgenic animal;  
KW gene therapy; drug screening; ds.  
XX  
OS Homo sapiens.  
XX  
PN DE19930512-Al.  
XX  
PD 11-JAN-2001.  
XX  
PF 05-JUL-1999; 99DE-1030512.  
XX  
PR 05-JUL-1999; 99DE-1030512.  
XX  
PA (BRUE/) BRUESS M.  
PA (BOEN/) BOENISCH H.  
XX  
PI Bruess M, Boenisch H;  
XX  
DR WPI; 2001-169861/18.  
DR P-PSDB; AAB46838.  
XX  
PT New human N-formyl receptor gene, useful for diagnosis and treatment of  
PT disease  
XX  
PS Disclosure; Page 4; 6pp; German.  
XX  
CC This invention describes a novel human N-formyl receptor gene (I),  
CC including its 5' and 3' untranslated regions. The invention also  
CC describes (a) transcription factors, RNA polymerases, pharmaceuticals  
CC and chemicals that up- or down-regulate expression of (I); (b) mRNA (II),  
CC and its splice variants or isoforms, transcribed from (I); (c) cDNA  
CC derived from (II) or from genes without introns; (d) protein (III)  
CC derived, or produced, from (II), cDNA or (I); (e) antibodies or antisera  
CC directed against one or more epitopes of (III) or the entire protein; (f)  
CC systems (including eukaryotic cells, yeast cells, Xenopus Oocytes,  
CC Baculovirus systems, and bacterial expression systems) that express  
CC native or recombinant (III); (g) ligand binding studies and screening  
CC assays that use the native or recombinant receptor, or cells or membranes  
CC that contain it; (h) transgenic and knockout animals that express the  
CC receptor at altered level or not at all; (i) gene therapy method that  
CC involves the receptor or its gene, cDNA or mRNA; (j) (anti)sense  
CC oligonucleotides derived from (I); (k) diagnosis and treatment of  
CC diseases in which the receptor is (in)directly implicated; (l)  
CC development of new (or evaluation of known) pharmaceuticals, compounds,  
CC chemicals, and techniques; and (m) modified versions of the protein,  
CC gene, cDNA, and mRNA sequences. (I), also related nucleic acids,  
CC proteins, antibodies, ligands etc., are potentially useful for diagnosis  
CC and (gene) therapy of diseases, also for drug screening, identification  
CC of ligands and production of transgenic animals.  
XX  
SQ Sequence 3540 BP; 748 A; 1146 C; 976 G; 670 T; 0 other;

Query Match 1.7%; Score 18; DB 22; Length 3540;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 821 tctctgactcttctgttc 838  
Db 798 tctctgactcttctgttc 815



RESULT 23  
AAT39809/C  
ID AAT39809 standard; DNA; 4091 BP.  
XX  
XX  
AC AAT39809;  
XX  
XX 23-FEB-1998 (first entry)  
XX  
XX Mouse H74 gene.  
XX  
XX Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;  
KW cellular signalling element; cellular structural element; malignancy;  
KW protein identification; functional domain; protein screening;  
KW cellular signal transduction process; ss.  
XX  
OS Mus musculus.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 200..1525  
FT /\*tag= a  
FT /\*transl\_except= (pos: 1514..1516, aa: Glu)  
XX  
XX WO9631625-A1.  
PN  
XX  
XX 10-OCT-1996.  
XX  
XX  
XX 04-APR-1996; 96WO-US04454.  
PF  
XX  
XX 03-APR-1996; 96US-0630915.  
PR  
XX 07-APR-1995; 95US-0417872.  
PR  
XX (CYTO-) CYTOGEN CORP.  
PA (UYN-) UNIV NORTH CAROLINA.  
XX  
XX Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;  
PI WPI; 1996-465045/46.  
XX P-PSDB; AAW05410.  
DR  
XX  
XX Identifying polypeptide(s) having specific functional domain (esp.  
PT SH3 domain) - comprises detecting selective binding to recognition  
PT unit, regardless of sequence homology  
XX  
XX  
PS Claim 102; Fig 46; 174pp; English.  
XX  
XX AAT39804-T39810 represent human and mouse genes encoding Src-homology  
CC region 3 (SH3) domain containing proteins that can be used in the method  
CC of the invention. SH3 domain containing proteins play a role in  
CC signalling and structural elements of cells. The method of the invention  
CC is for identifying polypeptides containing functional domains of  
CC interest (especially SH3 domains). The method comprises contacting a  
CC multivalent recognition unit (RU) complex with a number of peptides and  
CC identifying polypeptides having a selective binding affinity for the RU  
CC complex. The method is based on functional similarities and does not rely  
CC on sequence similarities. Prior methods only gave limited success for  
CC identifying proteins which contain an SH3 domain due to the minimal  
CC sequence homology among known SH3 proteins. It has been found that small  
CC peptide RUs in multivalent form have reduced specificity for a given  
CC functional domain compared to monomer RUs. Multivalent RU complexes are  
CC particularly suited to screening for polypeptides containing functional  
CC domains that are similar to, but not identical in sequence to, the  
CC original target functional domain. The new method enables proteins having  
CC a common function to be identified. Identification of novel SH3 proteins  
CC will be useful for a better understanding of cell growth, malignancy,  
CC signal transduction processes, etc. New candidate drugs can be  
CC identified, and their specificities (e.g. pharmacological activities) can  
CC be assessed using the method of the invention.  
XX  
SQ Sequence 4091 BP; 945 A; 1232 C; 1116 G; 798 T; 0 other;

Query Match

1.7%; Score 18; DB 17; Length 4091;

Best Local Similarity 100.0%; Pred. NO. 28;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 939 acagggctgggaagaa 956  
|||||  
DB 2031 ACAGGGCTGGGAAGAA 2014  
RESULT 24  
AAC75782  
ID AAC75782 standard; cDNA; 5035 BP.  
XX  
XX AAC75782;  
XX  
XX 08-FEB-2001 (first entry)  
XX  
XX Human ORFX ORF1337 polynucleotide sequence SEQ ID NO:2673.  
XX  
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW immunosuppressant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200058473-A2.  
XX  
XX 05-OCT-2000.  
XX  
XX 31-MAR-2000; 2000WO-US08621.  
XX  
XX 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
XX 30-MAR-2000; 2000US-0540763.  
XX  
XX (CURA-) CURAGEN CORP.  
PA  
XX Shimkets RA, Leach M;  
PI  
XX WPI; 2000-602362/57.  
XX P-PSDB; AAB41573.  
XX  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
XX Claim 5; Page 1918-1921; 5507pp; English.  
XX  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; antiviral; antirheumatic;  
CC antiinflammatory; antibacterial; antianaemic. The sequences can be used for determining  
CC antithyroid; and antianaemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,

CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease: to enhance  
CC coagulation: to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 5035 BP; 1160 A; 1424 C; 1490 G; 961 T; 0 other;

Query Match 1.7%; Score 18; DB 21; Length 5035;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 atggggaaaagttaagga 473  
|||||  
Db 2871 atggggaaaagttaagga 2888

RESULT 25  
AAX22304  
ID AAX22304 standard; DNA; 28866 BP.  
XX  
AC AAX22304;  
XX  
DT 22-JUN-1999 (first entry)  
XX  
DE Human IL-1ra BAC contiguous DNA sequence 97.  
XX  
KW Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;  
KW interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;  
KW chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;  
KW growth factors; treatment; IL-1 receptor complex; BAC; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO906426-A1.  
XX  
PD 11-FEB-1999.  
XX  
PF 03-AUG-1998; 98WO-US16102.  
XX  
PR 02-JUL-1998; 98US-0091650.  
PR 04-AUG-1997; 97US-0054646.  
XX  
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.  
XX  
PI Pan Y;  
XX  
DR WPI; 1999-153692/13.  
XX  
PT New isolated nucleic acid encoding the new human cytokine Tango-77 -  
PT used to inhibit inflammation and to screen for specific modulators  
PS Example 5; Figure 4; 226pp; English.  
XX  
CC AAX02956-X03048 and AAX22301-X22304 are overlapping BAC genomic  
CC sequences containing alternatively spliced forms of human IL-1ra. Such  
CC fragments are used in the method of the invention which describes the  
CC isolation of a novel human TANGO-77 encoding nucleic acid and protein.  
CC Tango-77 is a member of the cytokine superfamily that is expected to  
CC inhibit inflammation by binding to the interleukin-1 receptor (IL-1R). It  
CC may also bind to a new receptor so could regulate other cellular  
CC processes associated with acute or chronic inflammation, e.g. asthma,  
CC chronic myelogenous leukaemia, rheumatoid arthritis, psoriasis and  
CC inflammatory bowel disease. It may also induce or suppress interleukins,  
CC cytokines and growth factors. Modulators of this protein are used to  
CC treat or prevent conditions associated with abnormal levels of  
CC inflammation, or activity of IL-1 or its receptor complex.  
XX  
SQ Sequence 28866 BP; 8044 A; 6348 C; 6742 G; 7729 T; 3 other;

Query Match 1.7%; Score 18; DB 20; Length 28866;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 717 ctggggattagaatgtgg 734  
|||||  
Db 12941 ctggggattagaatgtgg 12958

RESULT 26  
AAD02697  
ID AAD02697 standard; DNA; 160552 BP.  
XX  
AC AAD02697;  
XX  
DT 02-MAY-2001 (first entry)  
XX  
DE Human glycosyl sulfotransferase-4 (GST-4) genomic DNA.  
XX  
KW Human; glycosyl sulfotransferase-4; GST-4; immunosuppressive;  
KW therapy; selectin binding inhibitor; gene therapy; inflammation;  
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;  
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;  
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; anaemia;  
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; eczema;  
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;  
KW myocarditis; adult respiratory distress syndrome; psoriasis;  
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;  
KW chromosome 16q23.1; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT exon 32847..32922  
FT /\*tag= a  
FT /\*number= 1  
FT /\*label= 4a\_5U4  
FT 32923..35592  
FT /\*tag= b  
FT /\*cons\_splice= (5'site:NO, 3'site:YES)  
FT 35593..35674  
FT /\*tag= c  
FT /\*number= 2  
FT /\*label= 4a\_5U3  
FT 35675..45093  
FT /\*tag= d  
FT 45094..45185  
FT /\*tag= e  
FT /\*number= 3  
FT /\*label= 4a\_5U2  
FT 45186..46633  
FT /\*tag= f  
FT /\*cons\_splice= (5'site:NO, 3'site:NO)  
FT 46634..46700  
FT /\*tag= g  
FT /\*number= 4  
FT /\*label= 4a\_5U1  
FT 46701..47938  
FT /\*tag= h  
FT /\*cons\_splice= (5'site:YES, 3'site:NO)  
FT 47939..49746  
FT /\*tag= i  
FT /\*number= 5  
FT /\*note= "Includes 17 base pairs of 5'UTR, the ORF  
FT and all of 3'UTR"  
FT 47939..47955  
FT /\*tag= j  
FT /\*note= "Portion of 5' untranslated region (5'UTR)"  
FT 47956..49128  
FT /\*tag= k  
FT /\*product= "Human glycosyl transferase-4alpha  
FT (GST-4alpha)"  
FT

```

FT FT 3'UTR 49129..49746
FT FT /*tag= l
FT FT exon 83257..83347
FT FT /*tag= m
FT FT /label= 4a_502
FT FT intron 83348..96412
FT FT /*tag= n
FT FT /cons_splice= (5'site:NO, 3'site:NO)
FT FT 96413..96484
FT FT /*tag= o
FT FT /label= 4a_501
FT FT intron 96485..98456
FT FT /*tag= p
FT FT /cons_splice= (5'site:NO, 3'site:NO)
FT FT 98457..99968
FT FT /*tag= q
FT FT /note= "Includes 17 base pairs of 5'UTR, the ORF
FT FT and all of 3'UTR"
FT FT 98457..98473
FT FT /*tag= r
FT FT /note= "Portion of 5' untranslated region (5'UTR)"
FT FT 98474..99661
FT FT /*tag= s
FT FT /product= "Human glycosyl transferase-4beta
FT FT (GST-4beta)"
FT FT 99662..99968
FT FT /*tag= t
FT FT
FT PN WO200106015-A1.
FT XX
FT XX 25-JAN-2001.
FT XX
FT XX 19-JUL-2000; 2000WO-US19741.
FT PF
FT XX
FT XX 20-JUL-1999; 99US-0144694.
FT PR
FT XX 13-JUL-2000; 2000US-0593828.
FT XX
FT PA (REGC ) UNIV CALIFORNIA.
FT XX
FT XX Rosen SD, Lee JK, Hemmerich S;
FT XX
FT XX WPI; 2001-138471/14.
FT DR
FT XX P-PSDB; AAY72639, AAY72640.
FT XX
FT XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
FT PT diagnostic and therapeutic agent screening applications -
FT XX
FT PS Example 1; Page 62-104; 128pp; English.
FT XX
FT CC The present sequence is human glycosyl sulfotransferase-4 (GST-4) genomic
FT CC DNA encoding GST-4alpha and GST-4beta. GST-4 gene is found on
FT CC chromosome 16q23.1.
FT CC GST is a type 2 membrane protein useful for inhibiting a binding event
FT CC between a selectin and a selectin ligand, which comprises contacting the
FT CC selectin with a non-sulphated selectin ligand, GST and a small molecular
FT CC agent that inhibits the sulphation activity of GST. GST is also useful
FT CC in inhibiting a selectin mediated binding event. GST is useful in gene
FT CC therapy to treat disorders such as acute or chronic inflammation,
FT CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
FT CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
FT CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
FT CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
FT CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
FT CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
FT CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
FT CC bronchial asthma, hypersensitivity, rheumatic fever, and tissue rejection
FT CC during transplantation.
FT XX
FT XX Sequence 160552 BP; 40281 A; 37573 C; 38015 G; 44564 T; 119 other;

```

Query Match 1.7%; Score 18; DB 22; Length 160552;  
 Best Local Similarity 100.0%; Pred. No. 27;

```

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 387 catcctgctatttgcct 404
Db 97322 catcctgctatttgcct 97339
RESULT 27
AAA00012
ID AAA00012 standard; cDNA; 300 BP.
XX
AC AAA00012;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:3.
XX
KW Human; colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9958675-A2.
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 99WO-US10602.
XX
PR 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 15-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0103234.
PR 27-OCT-1998; 98US-0105677.
XX
PA (CHIR ) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Glese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
DR WPI; 2000-126369/11.
XX
PT Polynucleotide library used to determine cancerous states of mammalian
PT cells -
XX
PS Claim 1; Page 178; 1097pp; English.
XX
CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC libraries constructed from human colon cancer cell lines. The present
CC invention also describes a method of detecting differentially expressed
CC genes correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotide sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX
SQ Sequence 300 BP; 70 A; 52 C; 44 G; 105 T; 29 other;

```

Query Match 1.6%; Score 17; DB 21; Length 300;

Best Local Similarity 100.0%; Pred. No. 91;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 621 tgatactcaaacatga 637  
|||||  
Db 142 tgatactcaaacatga 158

RESULT 28  
AAZ98178  
ID AAZ98178 standard; cDNA; 432 BP.  
XX AC  
XX AAZ98178;  
XX DT  
XX 11-MAY-2000 (first entry)  
XX DE  
XX Human signal peptide containing protein HSPP-70 cDNA SEQ ID NO:204.  
XX KW  
XX Human; signal peptide-containing protein; HSPP; diagnosis; cancer;  
KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;  
KW antimicrobial; nontropic; neuroprotective; cardiovascular; hepatotropic;  
KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;  
KW reproductive disorder; developmental disorder; arteriosclerosis;  
KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;  
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;  
KW Parkinson's disease; Huntington's diseases; ovulatory defect;  
KW muscular dystrophy; ss.  
XX OS  
XX Homo sapiens.  
XX PN  
XX WO200000610-A2.  
XX PD  
XX 06-JAN-2000.  
XX XX  
XX 25-JUN-1999; 99WO-US14484.  
XX XX  
XX 26-JUN-1998; 98US-0090762.  
PR 31-JUL-1998; 98US-0094983.  
PR 01-OCT-1998; 98US-0102686.  
PR 11-DEC-1998; 98US-0112129.  
XX XX  
XX (INCY-) INCYTE PHARM INC.  
XX PI  
XX Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;  
PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;  
PI Bandman O;  
XX XX  
XX WPI; 2000-160673/14.  
DR P-PSDB; AAY87293.  
XX XX  
XX New human signal peptide-containing proteins useful in treatment,  
PT prevention and diagnosis of e.g. cancer, inflammation and  
PT cardiovascular disease -  
XX XX  
XX Claim 9; Page 291; 327pp; English.

AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the  
human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have  
anticancer, anti-inflammatory, antimicrobial, nontropic, hepatotropic,  
neuroprotective, cardiovascular and antiasthmatic activities, and can  
be used in gene therapy. HSPPs can be used to treat or prevent disorders  
associated with decreased activity or function of HSPP. Antagonists of  
HSPP are used to treat or prevent disorders associated with increased  
activity or function of HSPP. Such diseases include cell proliferation  
(including cancer), inflammation, cardiovascular, neurological,  
reproductive or developmental disorders, (e.g. arteriosclerosis,  
cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,  
asthma, Crohn's disease, microbial or other infections, congestive or  
ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's  
diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP  
nucleic acids can be used for the recombinant production of HSPP, for  
detecting HSPP in standard hybridisation and amplification assays (for  
diagnosis and monitoring), in gene therapy, as antisense.

CC triplex-forming or ribozyme therapeutics, for detecting related sequences  
or genetic variations, and for chromosomal mapping. HSPP are also used to  
CC raise specific antibodies (Ab) and to screen for agonists and  
CC antagonists (potential therapeutic agents). Ab are used to diagnose, or  
CC monitor, HSPP-related diseases (in usual immunosays), as therapeutic  
CC antagonists, in competitive drug screens, and for purification of HSPP  
CC from natural sources.  
XX  
SQ Sequence 432 BP; 113 A; 102 C; 101 G; 116 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 432;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 942 gggctgggaagagaatg 958  
|||||  
Db 306 gggctgggaagagaatg 322

RESULT 29  
AAFL17622  
ID AAF17622 standard; cDNA; 451 BP.  
XX AC  
XX AAF17622;  
XX DT  
XX 13-MAR-2001 (first entry)  
XX DE  
XX Human breast cancer associated 13131 coding sequence.  
XX KW  
XX Human; breast cancer associated gene; vaccine; diagnosis; therapy; ss.  
XX OS  
XX Homo sapiens.  
XX PN  
XX WO200060076-A2.  
XX PD  
XX 12-OCT-2000.  
XX XX  
XX 15-FEB-2000; 2000WO-US05308.  
XX XX  
XX 02-APR-1999; 99US-0285480.  
PR 23-JUN-1999; 99US-0339338.  
PR 02-SEP-1999; 99US-0389881.  
PR 03-NOV-1999; 99US-0433826.  
XX XX  
XX (CORI-) CORIXA CORP.  
XX PI  
XX Yuqiu J, Dillon DC, Mitcham JL, Xu J, Harlocker SL;  
XX XX  
XX WPI; 2001-122627/13.  
XX XX  
XX An isolated polypeptide useful for the treatment and diagnosis of  
PT tumors e.g. breast cancer comprises at least an immunogenic portion of  
PT a breast tumor protein -  
XX XX  
XX Claim 6; Page 119; 238pp; English.

The present invention provides the coding sequences and some protein  
sequences of proteins associated with breast cancer in humans. These  
sequences can be used in the diagnosis and treatment of cancers,  
CC particularly breast tumours.  
XX  
SQ Sequence 451 BP; 138 A; 118 C; 101 G; 94 T; 0 other;

Query Match 1.6%; Score 17; DB 22; Length 451;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 gccaggatgctgctgca 108  
|||||  
Db 347 gccaggatgctgctgca 363

```
RESULT 30
AAA73713/c
ID AAA73713 standard; DNA; 468 BP.
XX
AC AAA73713;
XX
DT 07-DEC-2000 (first entry)
XX
DE 5' extension of PITSLRE protein kinase internal ribosome entry site.
XX
KW IRES; internal ribosome entry site; PITSLRE; human; gene therapy;
KW cancer; restenosis; p58; p110; protein kinase; ds.
XX
OS Homo sapiens.
XX
PN WO200044896-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-EP00643.
XX
PR 26-JAN-1999; 99EP-0200216.
XX
PA (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
XX
PI Cornelis S, Beyaert R;
XX
DR WPI; 2000-499331/44.
XX
PT Nucleic acids encoding internal ribosome entry sequences useful for
PT directing protein expression in gene therapy procedures -
XX
PS Claim 9; Page 33; 57pp; English.
XX
CC The PITSLRE protein kinase gene can be translated to give two isoforms,
CC p110 and p58. Transcription of p58 mRNA was found to start at an
CC internal ribosome entry site (IRES). The IRES element was found to be
CC cell cycle regulated, p58 is produced during the G2/M stage of the cell
CC cycle. The present sequence contains a 5' fragment of the PITSLRE
CC protein kinase IRES and nucleotides upstream of the IRES. The IRES
CC sequence and fragments of it may be used to induce cell cycle
CC dependent initiation of translation in eukaryotic cells. Vectors
CC containing the IRES may be used for the preparation of compositions for
CC the treatment of and/or prevention of a disease by gene therapy. Such
CC diseases may be cancer and restenosis.
XX
SQ Sequence 468 BP; 144 A; 100 C; 182 G; 42 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 468;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 cgctgctcctctgctg 42
| | | | | | | | | | |
Db 203 CGTGTCTCTCTGCTG 187

RESULT 31
AAA73714/c
ID AAA73714 standard; DNA; 660 BP.
XX
AC AAA73714;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human PITSLRE protein kinase gene IRES and upstream sequence.
XX
KW IRES; internal ribosome entry site; PITSLRE; human; gene therapy;
KW cancer; restenosis; p58; p110; protein kinase; ds.
XX
OS Homo sapiens.
```

```
XX
FH key Location/Qualifiers
FT misc_signal 439..660
FT /*tag= a
FT /label= IRES
FT /note= "Internal ribosome entry site"
XX
PN WO200044896-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-EP00643.
XX
PR 26-JAN-1999; 99EP-0200216.
XX
PA (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
XX
PI Cornelis S, Beyaert R;
XX
DR WPI; 2000-499331/44.
XX
PT Nucleic acids encoding internal ribosome entry sequences useful for
PT directing protein expression in gene therapy procedures -
XX
PS Claim 10; Page 34; 57pp; English.
XX
CC The PITSLRE protein kinase gene can be translated to give two isoforms,
CC p110 and p58. Transcription of p58 mRNA was found to start at an
CC internal ribosome entry site (IRES). The IRES element was found to be
CC cell cycle regulated, p58 is produced during the G2/M stage of the cell
CC cycle. The present sequence contains the PITSLRE protein kinase
CC IRES and upstream nucleotides. The IRES sequence and fragments of it
CC may be used to induce cell cycle dependent initiation of translation in
CC eukaryotic cells. Vectors containing the IRES may be used for the
CC preparation of compositions for the treatment of and/or prevention of
CC a disease by gene therapy. Such diseases may be cancer and restenosis.
XX
SQ Sequence 660 BP; 210 A; 128 C; 261 G; 61 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 660;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 cgctgctcctctgctg 42
| | | | | | | | | | |
Db 203 CGTGTCTCTCTGCTG 187

RESULT 32
AAAX40117/c
ID AAAX40117 standard; DNA; 677 BP.
XX
AC AAAX40117;
XX
DT 02-JUL-1999 (first entry)
XX
DE Gastric cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
```

PR 10-OCT-1997; 97US-0061599.  
 PR 10-OCT-1997; 97US-0061765.  
 PR 10-OCT-1997; 97US-0948705.  
 PR 11-OCT-1997; 97GB-0021697.  
 XX  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 XX  
 PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;  
 PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;  
 PI Tureci O;  
 XX  
 DR WPI; 1999-132448/11.  
 XX  
 PT New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 XX  
 PS Claim 67; Page 703; 787pp; English.  
 XX  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 XX  
 SQ Sequence 677 BP; 170 A; 161 C; 194 G; 145 T; 7 other;

Query Match 1.6%; Score 17; DB 20; Length 677;  
 Best Local Similarity 100.0%; Pred. No. 91;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 347 cctgttgaaactgtcc 363  
 |||||  
 Db 533 CCTGTGTTGAAACTGTCC 517

RESULT 33  
 AAX23298  
 ID AAX23298 standard; DNA; 701 BP.

AC AAX23298;  
 XX  
 DT 11-JUN-1999 (first entry)  
 XX  
 DE Human TRYI trypsinogen variant cDNA.

XX TRYI: trypsinogen; autocatalysis; cleavage; zymogenic precursor;  
 KW protease; peptide hormone; therapeutic protein; treatment; coagulation;  
 KW ss.

OS Homo sapiens.

XX WO9910503-A1.

PN 04-MAR-1999.

XX 12-AUG-1998; 98WO-EP05094.

XX 15-OCT-1997; 97EP-0117816.

PR 22-AUG-1997; 97EP-0114513.

XX (HOFF ) ROCHE DIAGNOSTICS GMBH.

XX

PI Bode W, Hopfner K, Huber R, Kopetzki E;  
 XX  
 DR WPI; 1999-204669/17.  
 DR P-PSDB; AAW93488.  
 XX  
 PT Autocatalytically cleavable zymogenic protease precursors - useful  
 PT for cleaving fusion proteins and for therapeutic uses  
 XX  
 PS Disclosure; Fig 1; 45pp; German.  
 XX  
 CC This invention describes a method where autocatalytically cleavable,  
 CC zymogenic precursors of a protease (in this invention trypsinogen),  
 CC have their natural cleavage site replaced by a non-natural,  
 CC autocatalytically cleavable site. Such proteases are reagents for  
 CC producing peptide hormones and other therapeutic proteins by cleavage  
 CC of their fusion proteins, and some also have therapeutic activity  
 CC themselves, e.g. thrombin for treatment of coagulation disorders. The  
 CC proteases are produced simply and in high yield without the need to  
 CC add another protease for cleavage, reducing costs and avoiding  
 CC contamination of the final enzyme.  
 XX  
 SQ Sequence 701 BP; 171 A; 194 C; 186 G; 150 T; 0 other;

Query Match 1.6%; Score 17; DB 20; Length 701;  
 Best Local Similarity 100.0%; Pred. No. 91;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 745 cctggagctctacaccaa 761

Db 625 cctggagctctacaccaa 641

RESULT 34

AAX78022  
 ID AAX78022 standard; DNA; 725 BP.

XX

AC AAX78022;

XX 19-AUG-1999 (first entry)

XX Chimeric serine protease FXT DNA.

XX Serine protease; chimeric; antithrombotic; modulator; drug design;  
 KW 3-D crystal structure; crystallization; haematopoietic cascade;  
 KW FXT; ss.

XX Synthetic.

XX EP927764-A2.

XX 07-JUL-1999.

XX 27-NOV-1998; 98EP-0122481.

XX 03-DEC-1997; 97EP-0121232.

XX (HOFF ) ROCHE DIAGNOSTICS GMBH.

XX Bode W, Engh R, Hopfner K, Huber R, Kopetzki E;

XX WPI; 1999-359878/31.

DR P-PSDB; AAY08894.

XX Chimeric serine protease comprising Factor X and Trypsin catalytic  
 PT domains, useful for identifying thrombolytic agents

XX Example 2; Fig 2; 23pp; German.

XX This invention describes a novel chimeric serine protease compound which  
 CC has antithrombotic activity and comprises two domains with a beta-sheet  
 CC structure. The first domain corresponds to the first domain of a first  
 CC serine protease and the second domain corresponds to the second domain

CC of a second serine protease. The products of the invention can be used  
 CC for identifying antithrombotic agents by determining whether an agent  
 CC modulates the activity of the serine protease. They can also be used to  
 CC identify agents through rational drug design using information based on  
 CC its 3-D crystal structure. The chimeric serine protease, is very good  
 CC for crystallization and for determining structural data and is also  
 CC useful for the identification of specific antithrombotic agents that  
 CC unlike prior art agents are extremely specific for only one factor in a  
 CC haematopoietic cascade.

SQ Sequence 725 BP; 172 A; 198 C; 216 G; 139 T; 0 other;

Query Match 1.6%; Score 17; DB 20; Length 725;  
 Best Local Similarity 100.0%; Pred. No. 91;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 cctggaggtctacacaa 761  
 |||||  
 Db 649 cctggaggtctacacaa 665

RESULT 35  
 AAN81633  
 ID AAN81633 standard; DNA; 741 BP.

XX AC AAN81633;

DT 07-NOV-1990 (first entry)

DE Human spleen trypsin III (trypsinogen III).

DE Human spleen plasminogen; trauma lesions; ss.

XX Homo sapiens.

XX FH Key Location/Qualifiers  
 FT CDS 1..741  
 FT /\*tag= a  
 FT /product=human spleen plasminogen III.

XX JP63160582-A.

XX PD 04-JUL-1988.

XX PF 25-DEC-1986; 86JP-0307770.

XX PR 25-DEC-1986; 86JP-0307770.

XX PA (SANY ) SANKYO KK.

XX DR WPI: 1988-224890/32.

XX DR P-PSDB; AAP81243.

PT Human spleen trypsin - used to treat lesions of trauma, without

PT hypersensitive allergic side effects.

PS Claim 8+9; Page 3; 9pp; Japanese.

XX Expression vectors E.coli LE 392 and YA 21 are preferable for mass  
 CC production, and animal cells or B.subtilis are suitable for the  
 CC production of an enzyme of similar activity to that of natural human  
 CC spleen trypsinogen. Culturing the recombinant cells produced the  
 CC desired trypsin as insoluble protein in inclusion bodies and the  
 CC trypsin can be isolated by lysing the cells by a suitable method. The  
 CC trypsin was then isolated and purified.  
 CC The product is used in the treatment of lesions or trauma, eg  
 CC burns, gangrene, abscesses, injury etc.

XX Sequence 741 BP; 160 A; 214 C; 200 G; 167 T; 0 other;

Query Match 1.6%; Score 17; DB 9; Length 741;

Best Local Similarity 100.0%; Pred. No. 91;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 cctggaggtctacacaa 761  
 |||||  
 Db 676 cctggaggtctacacaa 692

RESULT 36

AAT03999

ID AAT03999 standard; cDNA to mRNA; 744 BP.

XX AC AAT03999;

DT 19-MAR-1996 (first entry)

DE Human pancreatic trypsin III cDNA.

DE Pancreatic trypsin III; trypsinogen; human; hydrolysis; ds.

XX Homo sapiens.

XX FH Key Location/Qualifiers  
 FT CDS 1..744  
 FT /\*tag= a  
 FT /product= pancreatic\_trypsin\_III

XX JP07184655-A.

XX PD 25-JUL-1995.

XX PF 25-DEC-1986; 86JP-0307770.

XX PR 25-DEC-1986; 86JP-0307770.

XX PR 25-DEC-1986; 86JP-0311512.

XX (SANY ) SANKYO CO LTD.

XX WPI: 1995-287966/38.

XX P-PSDB; AAR87203.

PT Novel human pancreatic trypsin III - can be easily produced by  
 PT recombinant methods

PS Claim 2; Page 6-7; 11pp; Japanese.

XX AAT03999-T04001 are all human cDNA sequences which code for pancreatic  
 CC trypsin III (AAR87203), the sequences differ only in their stop codons.  
 CC The cDNA molecules can be used in the recombinant production of  
 CC trypsin which can be used as a drug to treat diseases wherein  
 CC trypsin production is impaired.

XX Sequence 744 BP; 162 A; 214 C; 200 G; 168 T; 0 other;

Query Match 1.6%; Score 17; DB 16; Length 744;

Best Local Similarity 100.0%; Pred. No. 91;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 cctggaggtctacacaa 761  
 |||||  
 Db 676 cctggaggtctacacaa 692

RESULT 37

AAT04000

ID AAT04000 standard; cDNA to mRNA; 744 BP.

XX AC AAT04000;

DT 19-MAR-1996 (first entry)

DE Human pancreatic trypsin III cDNA.

```
XX Pancreatic trypsin III; trypsinogen; human; hydrolysis; ds.
KW
XX
OS Homo sapiens.
XX
XX Location/Qualifiers
FT CDS 1..744
FT FT /*tag= a
FT FT /product= pancreatic_trypsin_III
XX
XX JP07184655-A.
XX
XX 25-JUL-1995.
XX
XX 25-DEC-1986; 86JP-0307770.
XX
XX 25-DEC-1986; 86JP-0307770.
XX
XX 25-DEC-1986; 86JP-0311512.
XX
XX (SANY ) SANKYO CO LTD.
XX
XX WPI; 1995-287966/38.
XX
XX P-PSDB; AAR82703.
XX
XX Novel human pancreatic trypsin III - can be easily produced by
XX recombinant methods
XX
XX Claim 3; Page 7-8; 11pp; Japanese.
XX
XX AAT03999-T04001 are all human cDNA sequences which code for pancreatic
XX trypsin III (AAR87203), the sequences differ only in their stop codons.
XX The cDNA molecules can be used in the recombinant production of
XX trypsin which can be used as a drug to treat diseases wherein
XX trypsin production is impaired.
XX
XX Sequence 744 BP; 161 A; 214 C; 201 G; 168 T; 0 other;

Query Match 1.6%; Score 17; DB 16; Length 744;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 cctggaggtctacaccaa 761
Db |||||||
676 cctggaggtctacaccaa 692

RESULT 38
AAT04001
ID AAT04001 standard; cDNA to mRNA; 744 BP.
XX
XX AAT04001;
XX
XX 19-MAR-1996 (first entry)
XX
XX Human pancreatic trypsin III cDNA.
XX
XX Pancreatic trypsin III; trypsinogen; human; hydrolysis; ds.
XX
XX Homo sapiens.
XX
XX Location/Qualifiers
FT CDS 1..744
FT FT /*tag= a
FT FT /product= pancreatic_trypsin_III
XX
XX JP07184655-A.
XX
XX 25-JUL-1995.
XX
XX 25-DEC-1986; 86JP-0307770.
XX
XX 25-DEC-1986; 86JP-0307770.
```

```
PR 25-DEC-1986; 86JP-0311512.
XX
XX (SANY ) SANKYO CO LTD.
XX
XX WPI; 1995-287966/38.
XX
XX Novel human pancreatic trypsin III - can be easily produced by
XX recombinant methods
XX
XX Claim 4; Page 9; 11pp; Japanese.
XX
XX AAT03999-T04001 are all human cDNA sequences which code for pancreatic
XX trypsin III (AAR87203), the sequences differ only in their stop codons.
XX The cDNA molecules can be used in the recombinant production of
XX trypsin which can be used as a drug to treat diseases wherein
XX trypsin production is impaired.
XX
XX Sequence 744 BP; 161 A; 214 C; 201 G; 168 T; 0 other;

Query Match 1.6%; Score 17; DB 16; Length 744;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 cctggaggtctacaccaa 761
Db |||||||
676 cctggaggtctacaccaa 692

RESULT 39
AAF15875/C
ID AAF15875 standard; cDNA; 771 BP.
XX
XX AAF15875;
XX
XX 13-MAR-2001 (first entry)
XX
XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:310.
XX
XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
XX neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
XX vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
XX antibacterial; gene therapy; neural; immune; reproductive; renal;
XX gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
XX wound; infectious disease; ss.
XX
XX Homo sapiens.
XX
XX WO20005174-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05988.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2000-587513/55.
XX
XX P-PSDB; AAB56672.
XX
XX Prostate cancer associated gene sequences, referred to as prostate
XX cancer antigens, useful for treatment, prevention, and diagnosis of
XX disorders such as prostate cancer -
XX
XX Claim 1; Page 821-822; 2338pp; English.
XX
XX AAF15566 to AAF16505 encode the human prostate cancer associated
XX proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
XX The prostate cancer antigens can have neuroprotective, cytostatic,
```



CC cardioactive, immunomodulatory, muscular, vulnery, gastrointestinal,  
CC nephrotropic, antineoplastic, gynaecological and antibacterial activities,  
CC and can be used in gene therapy. The prostate cancer antigen  
CC polynucleotides may be used for detection of prostate cancer, chromosome  
CC identification, as chromosome markers, and for numerous other diagnostic  
CC or research purposes. The prostate cancer antigens may be used to treat  
CC disorders such as neural, immune, muscular, reproductive,  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
CC AAB57303 represent sequences used in the exemplification of the present  
CC invention.  
XX  
SQ Sequence 771 BP; 249 A; 152 C; 147 G; 222 T; 1 other;

Query Match 1.6%; Score 17; DB 21; Length 771;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 348 cttgttgaaactgtcct 364  
|||||  
Db 449 CTTGTTGAAACTGTCTCT 433

RESULT 40  
AAV24548  
ID AAV24548 standard; cDNA; 790 BP.

XX AC AAV24548;

DT 16-SEP-1998 (first entry)

DE Trypsinogen-like protein coding sequence.

KW Trypsinogen-like protein; human; ds.

XX OS Homo sapiens.

FH Key Location/Qualifiers  
FT CDS 1..723  
FT /\*tag= a  
FT /product= trypsinogen-like protein

XX JP10099080-A.  
XX  
XX 21-APR-1998.

PF 26-SEP-1996; 96JP-0273923.

XX 26-SEP-1996; 96JP-0273923.

XX (SHIS ) SHISEIDO CO LTD.

XX WPI; 1998-289873/26.

DR P-PSDB; AAW57740.

XX DNA coding for trypsinogen-like protein

PT  
XX  
XX  
PS Claim 1; Page 4-5; 7pp; Japanese.

CC This sequence represents the gene of the invention, and encodes a human  
CC trypsinogen-like protein.  
XX  
XX Sequence 790 BP; 183 A; 234 C; 205 G; 168 T; 0 other;

Query Match 1.6%; Score 17; DB 19; Length 790;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 cctggagtctacaccaa 761  
|||||  
Db 655 cctggagtctacaccaa 671

RESULT 41  
AAT15535  
ID AAT15535 standard; cDNA to mRNA; 807 BP.

XX AC AAT15535;

DT 24-JUL-1997 (revised)

DT 11-NOV-1996 (first entry)

XX DNA encoding rat caldecrin.

XX caldecrin; rat; human; calcium serum lowering agent;

KW recombinant production; preprosequence; ds.

XX Rattus sp.

FH Key Location/Qualifiers

FT misc\_feature 1..39

FT /\*tag= a

FT /note= "pre-pro-sequence"

FT misc\_feature 46..87

FT /\*tag= b

FT /note= "pro-sequence"

FT mat\_peptide 88..804

FT /\*tag= c

FT /product= mature\_caldecrin

XX WO9600287-A.

XX 04-JAN-1996.

XX 23-JUN-1995; 95WO-JP01268.

XX 08-MAR-1995; 95JP-0074676.

PR 24-JUN-1994; 94JP-0164898.

XX (CHUS ) CHUGAI SEIYAKU KK.

XX Akiyama M, Noikura T, Saheki T, Tomomura A;

PI WPI; 1996-068871/07.

DR P-PSDB; AAR90682.

XX DNA encoding rat and human caldecrin and related vectors - for prodn.

PT of caldecrin for use as serum calcium lowering agent

XX Claim 1; Page 23-25; 43pp; Japanese.

XX The DNA encodes the rat caldecrin preprosequence (see AAR90682).

CC Expression vectors contg. the DNA can be used for the recombinant

CC production of caldecrin from transformant host cells. The caldecrin is

CC used as a serum calcium lowering agent.

CC (Revised entry submitted to correct sequence analysis breakdown.)

XX Sequence 807 BP; 176 A; 233 C; 232 G; 166 T; 0 other;

Query Match 1.6%; Score 17; DB 17; Length 807;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 gctggccttggcaggctc 129  
|||||

Db 119 gctggccttggcaggctc 135  
|||||

RESULT 42

AAZ77512/c

ID AAZ77512 standard; cDNA; 954 BP.

XX AC AAZ77512;

```

XX 10-APR-2000 (first entry)
XX Human ovarian tumor cDNA library derived EST fragment 63.
DE Expressed sequence tag; EST; human; ovarian tumor; anticancer;
XX gene therapy; treatment; ss.
KW Homo sapiens.
OS
XX DE19817557-A1.
XX 21-OCT-1999.
XX
XX 09-APR-1998; 98DE-1017557.
XX
XX 09-APR-1998; 98DE-1017557.
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;
XX WPI; 1999-591920/51.
XX
XX New nucleic acid sequences expressed in ovarian, and some other, cancer
XX tissues, and derived polypeptides, for treatment of ovarian cancer and
XX identification of therapeutic agents -
XX
XX Claim 3; Page 189; 310pp; German.
XX
XX This invention describes novel nucleic acid (cDNA) sequences (A) which
XX have anticancer activity and are highly expressed in ovarian tumor
XX tissue (and some also in testis and breast cancer tissue). The products
XX of the invention can be used for gene therapy. (A) are used (i) for
XX recombinant expression of polypeptides (B) and (ii) to isolate complete
XX genes. (B) are used (i) to identify agents suitable for treatment of
XX ovarian cancer; (ii) directly for treating this form of cancer
XX (including expression from gene therapy vectors) and (iii) for generation
XX of specific antibodies. (A) are identified by assembling ESTs (expressed
XX sequence tags) from a particular tissue type before comparison of
XX expression patterns. This allows a significantly longer fragment of the
XX gene to be revealed, so should reduce the number of failures associated
XX with the fact that ESTs from different libraries may represent different
XX parts of the same unknown gene, distorting the estimated frequency of
XX occurrence in a particular tissue. AA277450-277572 represent the human
XX ovarian tumor cDNA library derived EST fragments described in the method
XX of the invention and encode the protein fragments represented in
XX AA76505-Y76638.
XX
XX Sequence 954 BP; 209 A; 236 C; 276 G; 233 T; 0 other;
XX
XX
XX Query Match 1.6%; Score 17; DB 20; Length 954;
XX Best Local Similarity 100.0%; Pred. No. 91;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 745 cctggagtcctacaccaa 761
XX |||||||
XX Db 167 CCTGGAGTCTACACCAA 151
XX
XX RESULT 43
XX AAV63911
XX ID AAV63911 standard; DNA; 1045 BP.
XX
XX AC AAV63911;
XX
XX 20-JAN-1999 (first entry)
XX
XX Helicobacter pylori amIF HP1238 open reading frame.
XX
XX Helicobacter pylori; amIE; aliphatic amidase; gastritis; dyspepsia;
XX chronic gastroduodenal disorder; peptic ulcer; porcine gastric ulcer;

```

```

KW detection; infection; PCR primer; ss.
XX Helicobacter pylori.
XX
XX WO9844094-A2.
XX
XX 08-OCT-1998.
XX
XX 27-MAR-1998; 98WO-EP01824.
XX
XX 28-MAR-1997; 97US-0041745.
XX
XX (INRM ) INST NAT SANTE & RECH MEDICALE.
XX (INSP ) INST PASTEUR.
XX
XX De Reuse H, Labigne A, Skouloubris S;
XX WPI; 1998-557106/47.
XX
XX New Helicobacter aliphatic amidase AmIE polypeptides and their
XX encoding sequences - used in diagnosis, treatment and prevention of
XX Helicobacter sp. infections in humans and animals
XX
XX Example 17; Fig 12; 51pp; English.
XX
XX The present sequence represents a Helicobacter pylori amIF open reading
XX frame from an example from the present invention. The present invention
XX describes H. pylori aliphatic amidase AmIE. AmIE polypeptides
XX and antibodies form immunogenic, pharmaceutical and therapeutic
XX compositions which can be used in methods for detecting, treating or
XX preventing Helicobacter sp. (particularly H. pylori and H. heilmannii)
XX infections, especially chronic gastroduodenal disorders like gastritis,
XX dyspepsia and peptic ulcers in man, and porcine gastric ulcers in pigs.
XX They are administered in immunologically/pharmaceutically effective
XX amounts by an oral, intradermal, intramuscular, intravenous or mucosal
XX route to a patient. A detection kit for a Helicobacter infection in
XX comprises any AmIE polynucleotides with any of the vectors given in the
XX present invention. The polypeptides are used to in a process to degrade
XX acrylamide, acetamide, propionamide and isobutyramide. The polypeptides
XX are also used to screen for active substrates that inhibit Helicobacter
XX sp. amidase activity. The polynucleotides encoding the aliphatic amidase
XX AmIE polypeptides are the first to be characterized at the molecular
XX level. Methods involving such polypeptides are preferred to urease-based
XX methods because of the presence of urea positive bacteria in porcine
XX gastrointestinal tracts.
XX
XX Sequence 1045 BP; 299 A; 217 C; 256 G; 273 T; 0 other;
XX
XX
XX Query Match 1.6%; Score 17; DB 19; Length 1045;
XX Best Local Similarity 100.0%; Pred. No. 91;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 713 taagctgggattagaa 729
XX |||||||
XX Db 803 taagctgggattagaa 819
XX
XX RESULT 44
XX AAX14433
XX ID AAX14433 standard; DNA; 1102 BP.
XX
XX AC AAX14433;
XX
XX 31-MAR-1999 (first entry)
XX
XX H. pylori GHPO 783 gene.
XX
XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
XX peptic ulcer disease; ss.
XX
XX Helicobacter pylori.
XX

```

```

FH Key          Location/Qualifiers
FT CDS          59..1062
XX              /*tag= a
XX
XX WO9843478-A1.
XX
XX 08-OCT-1998.
XX
XX 01-APR-1998; 98WO-US06371.
XX
XX 29-JUL-1997; 97US-0902615.
XX 01-APR-1997; 97US-0833457.
XX 24-JUN-1997; 97US-0881227.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
XX Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
XX WPI; 1998-542293/46.
XX P-PSDB; AAW98714.
XX
XX New isolated Helicobacter polynucleotides - used to develop products
XX for the diagnosis, prevention and treatment of Helicobacter
XX PT infections and gastrointestinal diseases
XX
XX Claim 1; Page 1507-1509; 2054pp; English.
XX
XX This sequence represents a polynucleotide of the invention. It was
XX isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
XX The polypeptides can be used for preventing or treating Helicobacter
XX infections, and gastroduodenal diseases associated with these
XX infections, including acute, chronic, and atrophic gastritis, and peptic
XX ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
XX for the production of antibodies. The products can also be used for
XX detection and diagnosis.
XX
XX Sequence 1102 BP; 316 A; 224 C; 265 G; 297 T; 0 other;

Query Match          1.6%; Score 17; DB 19; Length 1102;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 713 taagctgggattagaa 729
Db 860 taagctgggattagaa 876
|||||

RESULT 45
AAZ20365
ID AAZ20365 standard; DNA; 1263 BP.
XX
XX AC AAZ20365;
XX
XX 17-NOV-1999 (first entry)
XX
XX DE Gene encoding bacterial general essential protein gep311.
XX
XX KW General essential protein; pathogenic bacteria; pathogen; inhibitor;
XX bacterial growth; ds.
XX
XX OS Streptococcus pneumoniae.
XX
XX FH Key          Location/Qualifiers
XX FT CDS          99..1262
XX              /*tag= a
XX              /product= gep311
XX
XX WO9333871-A2.
XX
XX 08-JUL-1999.

PF 30-DEC-1998; 98WO-US27918.
XX
XX 31-DEC-1997; 97US-0070116.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Youngman P, Fritz C, Murphy C, Guzman L;
XX
XX WPI; 1999-430230/36.
XX P-PSDB; AAY22574.
XX
XX Streptococcus pneumoniae general essential protein genes and proteins,
XX useful for identification of antibacterial agents -
XX
XX Claim 2; Fig 18; 124pp; English.
XX
XX This sequence represents a Streptococcus pneumoniae general essential
XX protein (GEP) gene of the invention. The genes encoding the GEP
XX polypeptides are useful molecular tools for identifying similar genes in
XX pathogenic microorganisms, such as pathogenic strains of Bacillus. In
XX addition, the operons containing genes encoding GEP and the polypeptides
XX themselves, are useful targets for identifying compounds that are
XX inhibitors of the pathogens in which the GEP are expressed. Such
XX inhibitors are useful for inhibiting bacterial growth by being
XX bacteriostatic or bacteriocidal.
XX
XX Sequence 1263 BP; 407 A; 244 C; 275 G; 337 T; 0 other;

Query Match          1.6%; Score 17; DB 20; Length 1263;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 764 taatctactaccacaaa 780
Db 346 taatctactaccacaaa 362
|||||

Search completed: November 22, 2001, 01:54:36
Job time: 3045 sec
```

---

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2001, 00:22:11 ; Search time 71.47 Seconds  
(without alignments)  
3298.777 Million cell updates/sec

Title: US-09-854-844-1  
Perfect score: 1041  
Sequence: 1 atgggcctgtggtgtgc.....ttattacaattgaaatga 1041

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2.6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2.6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2.6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2.6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2.6/ptodata/2/ina/PTUS\_COMB.seq: \*  
6: /cgn2.6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18	1.7	1959	4 US-09-201-641-1	Sequence 1, Appl
C 2	18	1.7	2038	4 US-09-008-271A-18	Sequence 18, Appl
C 3	18	1.7	176373	3 US-09-128-155-17	Sequence 17, Appl
C 4	17	1.6	677	4 US-08-896-164-42	Sequence 42, Appl
C 5	17	1.6	725	3 US-09-197-801-12	Sequence 12, Appl
C 6	17	1.6	725	4 US-09-551-028-12	Sequence 12, Appl
C 7	17	1.6	1347	4 US-08-896-164-84	Sequence 84, Appl
C 8	17	1.6	2409	4 US-09-293-322C-8	Sequence 8, Appl
C 9	17	1.6	2418	3 US-08-669-286-4	Sequence 4, Appl
C 10	17	1.6	2418	4 US-09-469-253-4	Sequence 4, Appl
C 11	17	1.6	2418	4 US-09-642-146-4	Sequence 4, Appl
C 12	17	1.6	2484	4 US-09-293-322C-4	Sequence 4, Appl
C 13	17	1.6	2630	3 US-08-669-286-6	Sequence 6, Appl
C 14	17	1.6	2630	4 US-09-469-253-6	Sequence 6, Appl
C 15	17	1.6	2630	4 US-09-642-146-6	Sequence 6, Appl
C 16	17	1.6	4529	2 US-08-449-645A-16	Sequence 16, Appl
C 17	17	1.6	4529	4 US-08-702-367A-16	Sequence 16, Appl
C 18	17	1.6	4529	5 PCT-US95-04681-16	Sequence 16, Appl
C 19	16	1.5	229	4 US-08-943-731-211	Sequence 211, App
C 20	16	1.5	529	3 US-09-188-930-24	Sequence 24, Appl
C 21	16	1.5	529	3 US-09-188-930-200	Sequence 200, App
C 22	16	1.5	996	1 US-08-121-057-1	Sequence 1, Appl
C 23	16	1.5	996	2 US-08-509-187D-1	Sequence 1, Appl
C 24	16	1.5	996	2 US-09-121-396-1	Sequence 1, Appl
C 25	16	1.5	996	5 PCT-US93-09704A-1	Sequence 1, Appl
C 26	16	1.5	1059	4 US-09-276-531-74	Sequence 74, Appl
C 27	16	1.5	1073	1 US-08-356-405-8	Sequence 8, Appl

C 28	1.5	1095	2 US-08-180-524-2	Sequence 2, Appl
C 29	1.5	1095	2 US-08-975-166-2	Sequence 2, Appl
C 30	1.5	1236	4 US-09-117-121-33	Sequence 33, Appl
C 31	1.5	1454	1 US-08-467-155A-2	Sequence 2, Appl
C 32	1.5	1454	2 US-08-628-198-2	Sequence 2, Appl
C 33	1.5	1454	3 US-09-201-038-2	Sequence 2, Appl
C 34	1.5	1454	5 PCT-US96-07343-2	Sequence 2, Appl
C 35	1.5	1622	1 US-07-996-772A-3	Sequence 3, Appl
C 36	1.5	1622	3 US-08-446-822-3	Sequence 3, Appl
C 37	1.5	1622	5 PCT-US93-12586-3	Sequence 3, Appl
C 38	1.5	1800	2 US-08-579-940-6	Sequence 6, Appl
C 39	1.5	1834	2 US-08-904-031-2	Sequence 2, Appl
C 40	1.5	1972	1 US-08-463-048-1	Sequence 1, Appl
C 41	1.5	1972	1 US-08-463-229-1	Sequence 1, Appl
C 42	1.5	1972	2 US-08-302-891-1	Sequence 1, Appl
C 43	1.5	3807	1 US-08-357-598-1	Sequence 1, Appl
C 44	1.5	3807	2 US-09-003-289-1	Sequence 1, Appl
C 45	1.5	3807	5 PCT-US95-16435-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-09-201-641-1/c  
; Sequence 1, Application US/09201641A  
; Patent No. 6232530  
; GENERAL INFORMATION:  
; APPLICANT: Cunningham Jr, Francis X  
; APPLICANT: Dellapenna, Dean  
; TITLE OF INVENTION: Method for Regulating Carotenoid Biosynthesis in  
; TITLE OF INVENTION: Marigolds  
; FILE REFERENCE: Quest 41-162  
; CURRENT APPLICATION NUMBER: US/09/201,641A  
; CURRENT FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 1959  
; TYPE: DNA  
; ORGANISM: Tagetes erecta  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (304)..(1836)  
; OTHER INFORMATION: beta-cyclase  
US-09-201-641-1

Query Match 1.7%; Score 18; DB 4; Length 1959;  
Best Local Similarity 100.0%; Pred. No. 7.9;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 997 gagccactgtaccctg 1014  
|||||  
DB 622 GAGCCACTGCTAACCTG 605

RESULT 2  
US-09-008-271A-18  
; Sequence 18, Application US/09008271A  
; Patent No. 6203979  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; Hillman, Jennifer L.  
; Yue, Henry  
; Guegler, Karl J.  
; Corley, Neil C.  
; Tang, Tom Y.  
; Shah, Purvi  
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.

```

; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheila
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2038 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLNNOT13
; CLONE: 1337018
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-008-271A-18

Query Match 1.7%; Score 18; DB 4; Length 2038;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 115 tggccttgccaggtcagc 132
|||||
Db 839 TGGCCTTGGCAGGTGTCAGC 856

RESULT 3
US-09-128-155-17
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pap, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(176373)
; OTHER INFORMATION: n .. A,T,C or G
US-09-128-155-17

; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheila
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2038 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLNNOT13
; CLONE: 1337018
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-008-271A-18

Query Match 1.7%; Score 18; DB 4; Length 2038;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 115 tggccttgccaggtcagc 132
|||||
Db 839 TGGCCTTGGCAGGTGTCAGC 856

RESULT 3
US-09-128-155-17
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pap, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(176373)
; OTHER INFORMATION: n .. A,T,C or G
US-09-128-155-17

; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheila
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2038 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLNNOT13
; CLONE: 1337018
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-008-271A-18

Query Match 1.7%; Score 18; DB 3; Length 176373;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 717 ctggggattagaatgtgg 734
|||||
Db 160448 ctggggattagaatgtgg 160465

RESULT 4
US-08-896-164-42/c
; Sequence 42, Application US/08896164
; Patent No. 6218521
; GENERAL INFORMATION:
; APPLICANT: OBATA, Yuichi
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED
; TITLE OF INVENTION: WITH GASTRIC CANCER AND METHODS FOR
; TITLE OF INVENTION: DIAGNOSING AND TREATING GASTRIC CANCER
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felife & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/896,164
; FILING DATE: July 17, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6218521man D. Hanson
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5499 - JEL/NDH/SLH
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 677 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-896-164-42

Query Match 1.6%; Score 17; DB 4; Length 677;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 347 ccttggtagaaactgtcc 363
|||||
Db 533 CCTTGTGAAACTGTCC 517

RESULT 5
US-09-197-801-12
; Sequence 12, Application US/09197801B
; Patent No. 6159722
; GENERAL INFORMATION:
; APPLICANT: Kopetzki, Erhard
; APPLICANT: Hopfner, Karl-Peter
; APPLICANT: Engh, Richard
; APPLICANT: Bode, Wolfram
; APPLICANT: Huber, Robert
; TITLE OF INVENTION: Chimeric Serine Proteases
; FILE REFERENCE: 20119
; CURRENT APPLICATION NUMBER: US/09/197,801B
```

; CURRENT FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 725  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-197-801-12

Query Match 1.6%; Score 17; DB 3; Length 725;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 cctggaggtctacaccaa 761  
|||||  
DB 649 cctggaggtctacaccaa 665

RESULT 6  
US-09-551-028-12  
; Sequence 12, Application US/09551028  
; Patent No. 6171842  
; GENERAL INFORMATION:  
; APPLICANT: Kopetzki, Erhard  
; APPLICANT: Hopfner, Karl-Peter  
; APPLICANT: Eng, Richard  
; APPLICANT: Bode, Wolfram  
; APPLICANT: Huber, Robert  
; TITLE OF INVENTION: Chimeric Serine Proteases  
; FILE REFERENCE: 20119  
; CURRENT APPLICATION NUMBER: US/09/551,028  
; PRIOR FILING DATE: 2000-04-17  
; PRIOR APPLICATION NUMBER: US/09/197,801  
; PRIOR FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 725  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-551-028-12

Query Match 1.6%; Score 17; DB 4; Length 725;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 cctggaggtctacaccaa 761  
|||||  
DB 649 cctggaggtctacaccaa 665

RESULT 7  
US-08-896-164-84/c  
; Sequence 84, Application US/08896164  
; Patent No. 6218521  
; GENERAL INFORMATION:  
; APPLICANT: OBATA, Yuichi  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED  
; WITH GASTRIC CANCER AND METHODS FOR  
; TITLE OF INVENTION: DIAGNOSING AND TREATING GASTRIC CANCER  
; NUMBER OF SEQUENCES: 87  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felife & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb  
; COMPUTER: IBM PS/2

; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/896,164  
; FILING DATE: July 17, 1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NO. 6218521man D. Hanson  
; REGISTRATION NUMBER: 30,946  
; REFERENCE/DOCKET NUMBER: LUD 5499 - JEL/NDH/SLH  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 688-9200  
; TELEFAX: (212) 838-3884  
; INFORMATION FOR SEQ ID NO: 84:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1347 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-896-164-84

Query Match 1.6%; Score 17; DB 4; Length 1347;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 cctgttgaaactgtcc 363  
|||||  
DB 533 CCTGTGAAACTGTCC 517

RESULT 8  
US-09-293-322C-8  
; Sequence 8, Application US/09293322C  
; Patent No. 6232110  
; GENERAL INFORMATION:  
; APPLICANT: Pallas, David C  
; APPLICANT: Du, Xianxing  
; TITLE OF INVENTION: Coding Sequence for Protein Phosphatase Methyltransferase,  
; Patent No. 6232110  
; TITLE OF INVENTION: Recombinant DNA Molecules and Methods  
; FILE REFERENCE: 105-97  
; CURRENT APPLICATION NUMBER: US/09/293,322C  
; CURRENT FILING DATE: 1999-04-16  
; PRIOR APPLICATION NUMBER: US 60/082,202  
; PRIOR FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 2409  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(2409)  
; OTHER INFORMATION: N is A, T, G or C.  
US-09-293-322C-8

Query Match 1.6%; Score 17; DB 4; Length 2409;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 916 gtagctgaagctgttc 932  
|||||  
DB 1138 gttagctgaagctgttc 1154

RESULT 9  
US-08-669-286-4/c  
; Sequence 4, Application US/08669286  
; Patent No. 6130060  
; GENERAL INFORMATION:

APPLICANT: NAKAMURA, SEIJI  
APPLICANT: SAKURAI, TAKASHI  
APPLICANT: NEZU, JUNI-ICHI  
TITLE OF INVENTION: GENE ENCODING ADSEVERIN  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/669,286  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: MURPHY JR., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-110P  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2418 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 27..2171  
US-08-669-286-4

Query Match 1.6%; Score 17; DB 3; Length 2418;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 347 cctgtgtgaactgtcc 363  
|||||  
Db 513 CCTGTGTAACGTCC 497

RESULT 10  
US-09-469-253-4/c  
Sequence 4, Application US/09469253  
Patent No. 6184352  
GENERAL INFORMATION:  
APPLICANT: NAKAMURA, SEIJI  
APPLICANT: SAKURAI, TAKASHI  
APPLICANT: NEZU, JUNI-ICHI  
TITLE OF INVENTION: GENE ENCODING ADSEVERIN  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
\* CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/469,253  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/669,286  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: MURPHY JR., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-110P  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2418 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 27..2171  
US-09-469-253-4

Query Match 1.6%; Score 17; DB 4; Length 2418;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 347 cctgtgtgaactgtcc 363  
|||||  
Db 513 CCTGTGTAACGTCC 497

RESULT 11  
US-09-642-146-4/c  
Sequence 4, Application US/09642146  
Patent No. 6271353  
GENERAL INFORMATION:  
APPLICANT: NAKAMURA, SEIJI  
APPLICANT: SAKURAI, TAKASHI  
APPLICANT: NEZU, JUNI-ICHI  
TITLE OF INVENTION: GENE ENCODING ADSEVERIN  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/642,146  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/669,286  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: MURPHY JR., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-110P  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:



; LENGTH: 2418 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 27..2171  
US-09-642-146-4

Query Match 1.6%; Score 17; DB 4; Length 2418;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 cctgttgaaactgtcc 363  
DB 513 CCTGTGAAACTGTCC 497

## RESULT 12

US-09-293-322C-4

; Sequence 4, Application US/09293322C

; Patent No. 6232110

; GENERAL INFORMATION:

; APPLICANT: Pallas, David C

; APPLICANT: Du, Xianxing

; TITLE OF INVENTION: Coding Sequence for Protein Phosphatase Methyltransferase,

; Patent No. 6232110

; TITLE OF INVENTION: Recombinant DNA Molecules and Methods

; FILE REFERENCE: 105-97

; CURRENT APPLICATION NUMBER: US/09/293,322C

; CURRENT FILING DATE: 1999-04-16

; PRIOR APPLICATION NUMBER: US 60/082,202

; PRIOR FILING DATE: 1998-04-17

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 2484

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (100)..(1257)

US-09-293-322C-4

## Query Match

1.6%; Score 17; DB 4; Length 2484;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 916 gtacgtgaagctgttc 932  
DB 1174 gtacgtgaagctgttc 1190

## RESULT 13

US-08-669-286-6/c

; Sequence 6, Application US/08669286

; Patent No. 6130060

; GENERAL INFORMATION:

; APPLICANT: NAKAMURA, SEIJI

; APPLICANT: SAKURAI, TAKASHI

; APPLICANT: NEZU, JUNI-ICHI

; TITLE OF INVENTION: GENE ENCODING ADSEVERIN

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP

; STREET: P.O. Box 747

; CITY: Falls Church

; STATE: VA

; COUNTRY: USA

; ZIP: 22040-0747

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/669,286  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURPHY Jr., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 230-110P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2630 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 79..2223  
US-08-669-286-6

- Query Match 1.6%; Score 17; DB 3; Length 2630;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 cctgttgaaactgtcc 363

DB 565 CCTGTGAAACTGTCC 549

## RESULT 14

US-09-469-253-6/c

; Sequence 6, Application US/09469253

; Patent No. 6184352

; GENERAL INFORMATION:

; APPLICANT: NAKAMURA, SEIJI

; APPLICANT: SAKURAI, TAKASHI

; APPLICANT: NEZU, JUNI-ICHI

; TITLE OF INVENTION: GENE ENCODING ADSEVERIN

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP

; STREET: P.O. Box 747

; CITY: Falls Church

; STATE: VA

; COUNTRY: USA

; ZIP: 22040-0747

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/469,253

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/669,286

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: MURPHY Jr., Gerald M.

; REGISTRATION NUMBER: 28,977

; REFERENCE/DOCKET NUMBER: 230-110P

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2630 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 79..2223  
US-09-469-253-6

Query Match 1.6%; Score 17; DB 4; Length 2630;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 ccttgtgaactgtcc 363  
|||||  
DB 565 CCTGTGAAACGTCC 549

RESULT 15  
US-09-642-146-6/c  
Sequence 6, Application US/09642146  
Patent No. 6271353  
GENERAL INFORMATION:  
APPLICANT: NAKAMURA, SEIJI  
APPLICANT: SAKURAI, TAKASHI  
APPLICANT: NEZU, JUN-ICHI  
TITLE OF INVENTION: GENE ENCODING ADSEVERIN  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/642,146  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/669,286  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: MURPHY Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-110P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2630 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 79..2223  
US-09-642-146-6

Query Match 1.6%; Score 17; DB 4; Length 2630;

Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 347 ccttgtgaactgtcc 363  
|||||  
DB 565 CCTGTGAAACGTCC 549

RESULT 16  
US-08-449-645A-16/c  
Sequence 16, Application US/08449645A  
Patent No. 5981245  
GENERAL INFORMATION:  
APPLICANT: Fox, Gary M.  
TITLE OF INVENTION: EPH-Like Receptor Protein Tyrosine  
TITLE OF INVENTION: Kinases  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Patent Operations/RBW  
STREET: 1840 Dehavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449,645A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-287  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4529 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 186...3182  
US-08-449-645A-16

Query Match 1.6%; Score 17; DB 2; Length 4529;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 tgatactgacagcagca 201  
|||||  
DB 1494 TGATACTGACAGCAGCA 1478

RESULT 17  
US-08-702-367A-16/c  
Sequence 16, Application US/08702367A  
Patent No. 5981246  
GENERAL INFORMATION:  
APPLICANT: Fox, Gary M.  
TITLE OF INVENTION: EPH-Like Receptor Protein Tyrosine  
TITLE OF INVENTION: Kinases  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Patent Operations/RBW  
STREET: 1840 Dehavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA

ZIP: 91320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/0702,367A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-287  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4529 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 186...3182  
US-08-702-367A-16

Query Match 1.6%; Score 17; DB 2; Length 4529;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 tgatactgacgacgca 201  
|||||  
DB 1494 TGATACTGACGACGCA 1478

RESULT 18  
PCT-US95-04681-16/c  
Sequence 16, Application PC/TUS9504681  
GENERAL INFORMATION:  
APPLICANT: Fox, Gary M.  
TITLE OF INVENTION: EPH-Like Receptor Protein Tyrosine  
TITLE OF INVENTION: Kinases  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Patent Operations/RBW  
STREET: 1840 Behavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04681  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-287  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4529 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 186...3182  
PCT-US95-04681-16

Query Match 1.6%; Score 17; DB 5; Length 4529;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 185 tgatactgacgacgca 201  
|||||  
DB 1494 TGATACTGACGACGCA 1478

RESULT 19  
US-08-943-731-211  
Sequence 211, Application US/08943731  
Patent No. 6265157  
GENERAL INFORMATION:  
APPLICANT: PROCKOP, DARWIN J.  
APPLICANT: SPOTILA, LORETTA D.  
APPLICANT: DELTAS, CONSTANTINOS D.  
APPLICANT: SEREDA, LARISSA  
APPLICANT: LARSON, ANDREA W.  
APPLICANT: PACK, MICHAEL  
APPLICANT: COLIGE, ALAIN  
APPLICANT: EARLY, JAMES  
APPLICANT: KORKKO, JARMO  
APPLICANT: ALA-KOKKO, LEENA, et al.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING  
TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES  
NUMBER OF SEQUENCES: 666  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.  
STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND  
STREET: FLR.  
CITY: PHILADELPHIA  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-7086  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,731  
FILING DATE: 03-OCT-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/212,322  
FILING DATE: 14-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/803,628  
FILING DATE: 03-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: DOYLE LEARY Ph.D., KATHRYN  
REGISTRATION NUMBER: 36,317  
REFERENCE/DOCKET NUMBER: 9598-27  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-965-1284  
TELEFAX: 215-567-2991  
TELEX: 831-494  
INFORMATION FOR SEQ ID NO: 211:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 229 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-943-731-211

Query Match 1.5%; Score 16; DB 4; Length 229;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 atgggccctgtgct 16
    |||||
Db 36 ATGGGCCCTGCTGGCT 51

RESULT 20
US-09-188-930-24
; Sequence 24, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 529
; TYPE: DNA
; ORGANISM: Rat
US-09-188-930-24

Query Match 1.5%; Score 16; DB 3; Length 529;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 928 gttgctgcatacagg 943
    |||||
Db 414 gttgctgcatacagg 429

RESULT 21
US-09-188-930-200
; Sequence 200, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 200
; LENGTH: 529
; TYPE: DNA
; ORGANISM: Rat
US-09-188-930-200

Query Match 1.5%; Score 16; DB 3; Length 529;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 928 gttgctgcatacagg 943
    |||||
Db 414 gttgctgcatacagg 429

RESULT 22
US-09-188-930-200
; Sequence 200, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 200
; LENGTH: 529
; TYPE: DNA
; ORGANISM: Rat
US-09-188-930-200

Query Match 1.5%; Score 16; DB 3; Length 529;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 928 gttgctgcatacagg 943
    |||||
Db 414 gttgctgcatacagg 429

RESULT 23
US-08-509-187D-1/c
; Sequence 1, Application US/08509187D
; Patent No. 5834283
; GENERAL INFORMATION:
; APPLICANT: Chang, Ta-Yuan and Chang, Catherine C.Y.
; TITLE OF INVENTION: ACYL Coenzyme A: Cholesterol Acyltransferase
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/509,187D
```

```
;
; FILING DATE: 31-JUL-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lamport Hammitte, Ann
; REGISTRATION NUMBER: 34,858
; REFERENCE/DOCKET NUMBER: DCI-033cpdv
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 996 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-509-187D-1

Query Match 1.5%; Score 16; DB 2; Length 996;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 caatgtaactactactac 774
|||||
Db 408 CAATGTAATCTACTAC 393

RESULT 24
US-09-121-396-1/c
; Sequence 1, Application US/09121396
; Patent No. 5968749
; GENERAL INFORMATION:
; APPLICANT: CHANG, TA-YUAN
; NAME: LAMPORT HAMMITTE, ANN.
; REGISTRATION NUMBER: 34,858
; TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
; TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/121,396
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/509,187
; FILING DATE: 07/31/95
; ATTORNEY/AGENT INFORMATION:
; NAME: LAMPORT HAMMITTE, ANN.
; REGISTRATION NUMBER: 34,858
; REFERENCE/DOCKET NUMBER: DCI-033CPDV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-2700
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 996 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

;
; FILING DATE: 31-JUL-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lamport Hammitte, Ann
; REGISTRATION NUMBER: 34,858
; REFERENCE/DOCKET NUMBER: DCI-033cpdv
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 996 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-09-121-396-1

Query Match 1.5%; Score 16; DB 2; Length 996;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 caatgtaactactactac 774
|||||
Db 408 CAATGTAATCTACTAC 393

RESULT 25
PCT-US93-09704A-1/c
; Sequence 1, Application PC/TUS9309704A
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
; TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
; NUMBER OF SEQUENCES: 9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/09704A
; FILING DATE: October 12, 1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. SER. NO. 959,950
; FILING DATE: October 14, 1992
; APPLICATION NUMBER: U.S. SER. NO. 121,057
; FILING DATE: September 10, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: LAMPORT HAMMITTE, ANN.
; REGISTRATION NUMBER: 34,858
; REFERENCE/DOCKET NUMBER: DCI-033CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-2700
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 996 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; PCT-US93-09704A-1

Query Match 1.5%; Score 16; DB 5; Length 996;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 caatgtaactactactac 774
|||||
Db 408 CAATGTAATCTACTAC 393

RESULT 26
US-09-276-531-74
; Sequence 74, Application US/09276531
; Patent No. 6183968
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Reddy, Roopa
; APPLICANT: Guegler, Karl J.
; APPLICANT: Baughn, Mariah R.
```

;; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
;; TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION  
;; NUMBER OF SEQUENCES: 134  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
;; STREET: 3174 PORTER DRIVE  
;; CITY: PALO ALTO  
;; STATE: CALIFORNIA  
;; COUNTRY: USA  
;; ZIP: 94304  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/276,531  
;; FILING DATE: March 27, 1998  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 60/079,677  
;; FILING DATE: March 27, 1998  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Lynn E. Murry, Ph.D.  
;; REGISTRATION NUMBER: 42,918  
;; REFERENCE/DOCKET NUMBER: PA-0008 US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (650) 855-0555  
;; TELEFAX: (650) 845-4166  
;; INFORMATION FOR SEQ ID NO: 74:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1059 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; IMMEDIATE SOURCE:  
;; LIBRARY: PANTUT01  
;; CLONE: 1516263  
US-09-276-531-74

Query Match 1.5%; Score 16; DB 4; Length 1059;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 656 attctgagggcctct 671  
|||||  
Db 208 ATTCTGGAGGCTCT 223

RESULT 27  
US-356-405-8/c  
; Sequence 8, Application US/08356405  
; Patent No. 5807691  
; GENERAL INFORMATION:  
; APPLICANT: Anlaiky, No. 5807691rdline  
; APPLICANT: Boschert, Ursula  
; APPLICANT: Hen, Rene  
; APPLICANT: Plassat, Jean-Luc  
; TITLE OF INVENTION: Polypeptides Having Serotonin Receptor  
; TITLE OF INVENTION: Activity (5HT5A), Nucleic Acids Coding for These  
; TITLE OF INVENTION: Polypeptides and Uses Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Arcola Road, 3C43  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426-0107  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/356,405  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: WO PCT/FR93/00650  
;; FILING DATE: 29-JUN-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: FR 92/08081  
;; FILING DATE: 01-JUL-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Smith, Julie K  
;; REGISTRATION NUMBER: 38,619  
;; REFERENCE/DOCKET NUMBER: EX92004-US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (610)454-3839  
;; TELEFAX: (610)454-3808  
;; INFORMATION FOR SEQ ID NO: 8:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1073 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
;; ORIGINAL SOURCE:  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..1071  
US-08-356-405-8

Query Match 1.5%; Score 16; DB 1; Length 1073;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 cgctgctctcttgcct 41  
|||||  
Db 845 CGCTGCTCTCTGCT 830

RESULT 28  
US-08-180-524-2/c  
; Sequence 2, Application US/08180524  
; Patent No. 5849537  
; GENERAL INFORMATION:  
; APPLICANT: Tripp, Matthew  
; APPLICANT: Lusk, Lance  
; APPLICANT: Rhodes, Thomas  
; APPLICANT: Huige, Nick  
; APPLICANT: Kot, Edward  
; APPLICANT: Chicoye, Etzer  
; APPLICANT: Barney, Michael C.  
; APPLICANT: Bower, Patricia A.  
; APPLICANT: Cronan, Charles L.

; TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS  
; TITLE OF INVENTION: IN YEAST  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Thad Kryshak, Quarles & Brady  
; STREET: 411 East Wisconsin Avenue  
; CITY: Milwaukee  
; STATE: Wisconsin  
; COUNTRY: USA  
; ZIP: 53202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MSWORD Version 5.0

;  
;  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/180,524  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/917,216  
; FILING DATE:  
; APPLICATION NUMBER: US 07/486,333  
; FILING DATE: 28-FEB-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/409,217  
; FILING DATE: 19-SEP-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kryshak, Thad  
; REGISTRATION NUMBER: 35,433  
; REFERENCE/DOCKET NUMBER: 66-005-9234-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (414) 277-5707  
; TELEFAX: (414) 277-5591  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1095 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Pseudopleuronectus americanus  
; US-08-180-524-2

Query Match 1.5%; Score 16; DB 2; Length 1095;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 742 cttcctggaggtctaca 757  
|||||  
Db 1067 CTCTGGAGTCTACA 1052

RESULT 29  
US-08-975-166-2/c  
; Sequence 2, Application US/08975166  
; Patent No. 5928877  
; GENERAL INFORMATION:  
; APPLICANT: Tripp, Matthew  
; APPLICANT: Lusk, Lance  
; APPLICANT: Rhodes, Thomas  
; APPLICANT: Huige, Nick  
; APPLICANT: Kot, Edward  
; APPLICANT: Chicoye, Etzer  
; APPLICANT: Barney, Michael C.  
; APPLICANT: Bower, Patricia A.  
; APPLICANT: Cronan, Charles L.  
; TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Thad Kryshak, Quarles & Brady  
; STREET: 411 East Wisconsin Avenue  
; CITY: Milwaukee  
; STATE: Wisconsin  
; COUNTRY: USA  
; ZIP: 53202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MSWORD Version 5.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/975,166

;  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/180,524  
; FILING DATE:  
; APPLICATION NUMBER: US/07/917,216  
; FILING DATE:  
; APPLICATION NUMBER: US 07/486,333  
; FILING DATE: 28-FEB-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/409,217  
; FILING DATE: 19-SEP-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kryshak, Thad  
; REGISTRATION NUMBER: 35,433  
; REFERENCE/DOCKET NUMBER: 66-005-9234-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (414) 277-5707  
; TELEFAX: (414) 277-5591  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1095 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Pseudopleuronectus americanus  
; US-08-975-166-2

Query Match 1.5%; Score 16; DB 2; Length 1095;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 742 cttcctggaggtctaca 757  
|||||  
Db 1067 CTCTGGAGTCTACA 1052

RESULT 30  
US-09-117-121-33/c  
; Sequence 33, Application US/09117121  
; Patent No. 6307020  
; GENERAL INFORMATION:  
; APPLICANT: Hew, Choy  
; APPLICANT: Gong, Zhiyuan  
; TITLE OF INVENTION: Intracellular Antifreeze Polypeptides  
; TITLE OF INVENTION: and Nucleic Acids  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/117,121  
; FILING DATE: 20-NOV-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/CA97/00062  
; FILING DATE: 30-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Kenneth A.

```
;
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 016252-001610US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1236 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..1236
; OTHER INFORMATION: /label=F2
; NAME/KEY: CDS
; LOCATION: 22..1041
; OTHER INFORMATION: /product="sAFP2"
; OTHER INFORMATION: /note="skin-type antifreeze"
; OTHER INFORMATION: polypeptide 2 (sAFP2)"
;
; US-09-117-121-33
;
; Query Match 1.5%: Score 16; DB 4; Length 1236;
; Best Local Similarity 100.0%; Pred. No. 84;
; Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 742 cttcttgagctctaca 757
; |
; Db 1206 CTTCTTGAGCTCTACA 1191
;
; RESULT 31
; US-08-467-155A-2
; Sequence 2, Application US/08467155A
; Patent No. 5736377
; GENERAL INFORMATION:
; APPLICANT: Band, Vimla
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED
; MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,155A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00398/100001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1454 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-628-198-2
;
; Query Match 1.5%: Score 16; DB 2; Length 1454;
; Best Local Similarity 100.0%; Pred. No. 85;
; Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 311 tccatcccaagtaccac 326
; |
; Db 431 TCCATCCCAAGTACCA 446
;
; RESULT 32
; US-08-628-198-2
; Sequence 2, Application US/08628198
; Patent No. 5843694
; GENERAL INFORMATION:
; APPLICANT: Band, Vimla
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED
; MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,198
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/467,155
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00398/100002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1454 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-628-198-2
;
; Query Match 1.5%: Score 16; DB 2; Length 1454;
; Best Local Similarity 100.0%; Pred. No. 85;
; Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 311 tccatcccaagtaccac 326
; |
; Db 431 TCCATCCCAAGTACCA 446
;
; RESULT 33
; US-09-201-038-2
; Sequence 2, Application US/09201038
; Patent No. 6153387
; GENERAL INFORMATION:
```



APPLICANT: Band, Vmla  
TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/201,038  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/628,198  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00398/100002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1454 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-201-038-2

Query Match 1.5%; Score 16; DB 3; Length 1454;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 tccatcccaagtacca 326  
|||||  
DB 431 TCCATCCCAAGTACCA 446

RESULT 34  
PCT-US96-07343-2  
SEQUENCE 2, Application PC/TUS9607343  
GENERAL INFORMATION:  
APPLICANT: New England Medical Center Hospitals, Inc.  
TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/07343  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/467,155  
FILING DATE: 06-JUN-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00398/100001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1454 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US96-07343-2

Query Match 1.5%; Score 16; DB 5; Length 1454;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 tccatcccaagtacca 326  
|||||  
DB 431 TCCATCCCAAGTACCA 446

RESULT 35  
US-07-996-772A-3  
SEQUENCE 3, Application US/07996772A  
PATENT NO. 5472866  
GENERAL INFORMATION:  
APPLICANT: Gerald, Christophe  
APPLICANT: Hartig, Paul R.  
APPLICANT: Branchek, Theresa A.  
APPLICANT: Weinshank, Richard L.  
TITLE OF INVENTION: DNA ENCODING 5-HT4A SEROTONIN  
RECEPTORS AND USES THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: COOPER & DUNHAM  
STREET: 30 ROCKEFELLER PLAZA  
CITY: NEW YORK  
STATE: NEW YORK  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/996,772A  
FILING DATE:

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, P., John  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 42667/JPW/TEP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOP UI  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1622 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown

MOLECULE TYPE: CDNA  
HYPOTHETICAL: N  
ANTI-SENSE: N  
ORIGINAL SOURCE: brain  
TISSUE TYPE: brain  
IMMEDIATE SOURCE:  
LIBRARY: rat brain  
CLONE: S10-95  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 50..1267  
OTHER INFORMATION:

US-07-996-772A-3

Query Match 1.5%; Score 16; DB 1; Length 1622;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 739 tctctctctgagctct 754  
|||||

Db 1438 TCCTCTCTGGAGTCT 1453

RESULT 36

US-08-446-822-3

Sequence 3, Application US/08446822

Patent No. 5766879

GENERAL INFORMATION:

APPLICANT: SYNAPTIC PHARMACEUTICAL CORPORATION

TITLE OF INVENTION: DNA ENCODING 5-HT<sub>4</sub> SEROTONIN RECEPTORS

TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: COOPER & DUNHAM

STREET: 30 ROCKEFELLER PLAZA

CITY: NEW YORK

STATE: NEW YORK

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/446,822

FILING DATE: June 1, 1995

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: White, P., John

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 42667-A-PCT-US/JPW/MAT

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1622 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: CDNA

HYPOTHETICAL: N

ANTI-SENSE: N

ORIGINAL SOURCE:

TISSUE TYPE: brain

IMMEDIATE SOURCE:

LIBRARY: rat brain

CLONE: S10-95

FEATURE:

NAME/KEY: CDS

LOCATION: 50..1267

OTHER INFORMATION:

US-08-446-822-3

Query Match 1.5%; Score 16; DB 1; Length 1622;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 739 tctctctctgagctct 754  
|||||

Db 1438 TCCTCTCTGGAGTCT 1453

RESULT 37

PCT-US93-12586-3

Sequence 3, Application PC/TUS9312586

GENERAL INFORMATION:

APPLICANT: SYNAPTIC PHARMACEUTICAL CORPORATION

TITLE OF INVENTION: DNA ENCODING 5-HT<sub>4</sub> SEROTONIN RECEPTORS

TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: COOPER & DUNHAM

STREET: 30 ROCKEFELLER PLAZA

CITY: NEW YORK

STATE: NEW YORK

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/12586

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: White, P., John

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 42667-A-PCT/JPW/TEP

TELEPHONE: (212) 977-9550

TELEFAX: (212) 664-0525

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1622 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: CDNA

HYPOTHETICAL: N

ANTI-SENSE: N

ORIGINAL SOURCE:

TISSUE TYPE: brain

IMMEDIATE SOURCE:

LIBRARY: rat brain

CLONE: S10-95

FEATURE:

NAME/KEY: CDS

LOCATION: 50..1267

OTHER INFORMATION:

PCT-US93-12586-3

Query Match 1.5%; Score 16; DB 5; Length 1622;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 739 tctctctctgagctct 754  
|||||

Db 1438 TCCTCTCTGGAGTCT 1453

```

RESULT 38
US-08-579-940-6/c
; Sequence 6, Application US/08579940
; Patent No. 5977315
; GENERAL INFORMATION:
; APPLICANT: Chatterjee, Malaya
; APPLICANT: Kohler, Heinz
; APPLICANT: Foon, Kenneth A.
; APPLICANT: Chatterjee, Sunil K.
; TITLE OF INVENTION: MURINE MONOCLONAL ANTI-IDIOTYPE ANTIBODY
; TITLE OF INVENTION: 3H1
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/579,940
; FILING DATE: 28-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H.
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 30414-20001.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(99..389, 746..784, 883..1203, 1325..1645)
; US-08-579-940-6

Query Match 1.5%; Score 16; DB 2; Length 1800;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 gctggtgatactcaaa 631
|||||
DB 627 GCTGGTGATCAAA 612

RESULT 39
US-08-904-031-2/c
; Sequence 2, Application US/08904031
; Patent No. 5948619
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN ZYGIN-1
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA

```

```

; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,031
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0353 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1834 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BEPINON01
; CLONE: 1424985
; US-08-904-031-2

Query Match 1.5%; Score 16; DB 2; Length 1834;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 cgctgctctcttctgct 41
|||||
DB 941 CGCTGCTCCTTCGTCT 926

RESULT 40
US-08-463-048-1/c
; Sequence 1, Application US/08463048
; Patent No. 5723760
; GENERAL INFORMATION:
; APPLICANT: STRITTMATTER, Gunter
; APPLICANT: MARTINI, No. 5723760bert
; TITLE OF INVENTION: FUNGUS-RESPONSIVE CHIMAERIC GENE
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,048
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR93/00700
; FILING DATE: 22-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm M

```

```

; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-068
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1972 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: potato cv. Datura
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..6
; OTHER INFORMATION: /note= "EcoRI site at position
; OTHER INFORMATION: 1-6."
; FEATURE:
; NAME/KEY: -
; LOCATION: 295..298
; OTHER INFORMATION: /note= "Sau3AI site at position
; OTHER INFORMATION: 295-298."
; FEATURE:
; NAME/KEY: -
; LOCATION: 643
; OTHER INFORMATION: /note= "5' end of putative TATA box
; OTHER INFORMATION: at position 643."
; FEATURE:
; NAME/KEY: -
; LOCATION: 751
; OTHER INFORMATION: /note= "ATG translation initiation
; OTHER INFORMATION: condon at position 751."
; FEATURE:
; NAME/KEY: -
; LOCATION: 1809
; OTHER INFORMATION: /note= "UGA stop codon at position
; OTHER INFORMATION: 1809."
; US-08-463-048-1
;
; Query Match 1.5%; Score 16; DB 1; Length 1972;
; Best Local Similarity 100.0%; Pred. No. 85;
; Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 375 cttcacttctgccatc 390
   |||||||
Db 765 CTTCACTTCTGCCATC 750

;
; RESULT 41
; US-08-463-229-1/c
; Sequence 1, Application US/08463229
; Patent No. 5750874
; GENERAL INFORMATION:
; APPLICANT: STRITTMATTER, Gunter
; APPLICANT: MARTINI, No. 5750874bert
; TITLE OF INVENTION: FUNGUS-RESPONSIVE CHIMAERIC GENE
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P. O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/463,229
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,891
; FILING DATE: 30-JAN-1995
; APPLICATION NUMBER: WO PCT/FR93/00700
; FILING DATE: 22-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm M
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-054
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1972 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: potato cv. Datura
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..6
; OTHER INFORMATION: /note= "EcoRI site at position
; OTHER INFORMATION: 1-6."
; FEATURE:
; NAME/KEY: -
; LOCATION: 295..298
; OTHER INFORMATION: /note= "Sau3AI site at position
; OTHER INFORMATION: 295-298."
; FEATURE:
; NAME/KEY: -
; LOCATION: 643
; OTHER INFORMATION: /note= "5' end of putative TATA box
; OTHER INFORMATION: at position 643."
; FEATURE:
; NAME/KEY: -
; LOCATION: 751
; OTHER INFORMATION: /note= "ATG translation initiation
; OTHER INFORMATION: condon at position 751."
; FEATURE:
; NAME/KEY: -
; LOCATION: 1809
; OTHER INFORMATION: /note= "UGA stop codon at position
; OTHER INFORMATION: 1809."
; US-08-463-229-1
;
; Query Match 1.5%; Score 16; DB 1; Length 1972;
; Best Local Similarity 100.0%; Pred. No. 85;
; Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 375 cttcacttctgccatc 390
   |||||||
Db 765 CTTCACTTCTGCCATC 750

;
; RESULT 42
; US-08-302-891-1/c
; Sequence 1, Application US/08302891
; Patent No. 5859332
; GENERAL INFORMATION:
; APPLICANT: STRITTMATTER, Gunter
; APPLICANT: MARTINI, No. 5859332bert
; TITLE OF INVENTION: FUNGUS-RESPONSIVE CHIMAERIC GENE
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P. O. Box 1404

```

;/ CITY: Alexandria  
;/ STATE: Virginia  
;/ COUNTRY: United States  
;/ ZIP: 22313-1404  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: Floppy disk  
;/ COMPUTER: IBM PC compatible  
;/ OPERATING SYSTEM: PC-DOS/MS-DOS  
;/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/08/302,891  
;/ FILING DATE: 30-JAN-1995  
;/ CLASSIFICATION: 800  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: WO PCT/FR93/00700  
;/ FILING DATE: 22-MAR-1993  
;/ NAME: McGowan, Malcolm M  
;/ ATTORNEY/AGENT INFORMATION:  
;/ REGISTRATION NUMBER: 39,300  
;/ REFERENCE/DOCKET NUMBER: 010830-054  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: (703) 836-6620  
;/ TELEFAX: (703) 836-2021  
;/ INFORMATION FOR SEQ ID NO: 1:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 1972 base pairs  
;/ TYPE: nucleic acid  
;/ STRANDEDNESS: double  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: DNA (genomic)  
;/ ORIGINAL SOURCE:  
;/ ORGANISM: potato cv. Datura  
;/ FEATURE:  
;/ NAME/KEY: -  
;/ LOCATION: 1..6  
;/ OTHER INFORMATION: /note= "EcoRI site at position  
;/ OTHER INFORMATION: 1-6."  
;/ FEATURE:  
;/ NAME/KEY: -  
;/ LOCATION: 295..298  
;/ OTHER INFORMATION: /note= "Sau3AI site at position  
;/ OTHER INFORMATION: 295-298."  
;/ FEATURE:  
;/ NAME/KEY: -  
;/ LOCATION: 643  
;/ OTHER INFORMATION: /note= "5' end of putative TATA box  
;/ OTHER INFORMATION: at position 643."  
;/ FEATURE:  
;/ NAME/KEY: -  
;/ LOCATION: 751  
;/ OTHER INFORMATION: /note= "ATG translation initiation  
;/ OTHER INFORMATION: condon at position 751."  
;/ FEATURE:  
;/ NAME/KEY: -  
;/ LOCATION: 1809  
;/ OTHER INFORMATION: /note= "UGA stop condon at position  
;/ OTHER INFORMATION: 1809."  
;/ US-08-302-891-1

Query Match 1.5%; Score 16; DB 2; Length 1972;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 cttcaactctgtccatc 390  
|||||  
DB 765 cttcactctgtccatc 750

RESULT 43  
US-08-357-598-1/c  
; Sequence 1, Application US/08357598  
; Patent No. 5705625

;/ GENERAL INFORMATION:  
;/ APPLICANT: Civin, Curt I.  
;/ APPLICANT: Small, Donald  
;/ TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3  
;/ NUMBER OF SEQUENCES: 12  
;/ CORRESPONDENCE ADDRESS:  
;/ ADDRESSEE: Fish & Richardson P.C.  
;/ STREET: 4225 Executive Square, Suite 1400  
;/ CITY: La Jolla  
;/ STATE: CA  
;/ COUNTRY: USA  
;/ ZIP: 92037  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: Floppy disk  
;/ COMPUTER: IBM PC compatible  
;/ OPERATING SYSTEM: PC-DOS/MS-DOS  
;/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/08/357,598  
;/ FILING DATE: 15-DEC-1994  
;/ CLASSIFICATION: 435  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Haile, Lisa A.  
;/ REGISTRATION NUMBER: 38,347  
;/ REFERENCE/DOCKET NUMBER: 07265/033001  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: 619/678-5070  
;/ TELEFAX: 619/678-5099  
;/ INFORMATION FOR SEQ ID NO: 1:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 3807 base pairs  
;/ TYPE: nucleic acid  
;/ STRANDEDNESS: single  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: DNA  
;/ US-08-357-598-1

Query Match 1.5%; Score 16; DB 1; Length 3807;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 gcgcgtggccttggcag 126  
|||||  
DB 352 GCGCTGCGCTTGGCAG 337

RESULT 44  
US-09-003-289-1/c  
; Sequence 1, Application US/09003289  
; Patent No. 5916792  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Civin, Curt I.  
;/ APPLICANT: Small, Donald  
;/ TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3  
;/ NUMBER OF SEQUENCES: 12  
;/ CORRESPONDENCE ADDRESS:  
;/ ADDRESSEE: Fish & Richardson P.C.  
;/ STREET: 4225 Executive Square, Suite 1400  
;/ CITY: La Jolla  
;/ STATE: CA  
;/ COUNTRY: USA  
;/ ZIP: 92037  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: Floppy disk  
;/ COMPUTER: IBM PC compatible  
;/ OPERATING SYSTEM: PC-DOS/MS-DOS  
;/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/09/003,289  
;/ FILING DATE:  
;/ CLASSIFICATION:  
;/ PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/357,598  
FILING DATE: 15-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/033001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3807 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-09-003-289-1

Query Match 1.5%; Score 16; DB 2; Length 3807;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 gcgctggccttggcag 126  
|||||  
Db 352 GCGCTGGCCTTGGCAG 337

RESULT 45  
PCT-US95-16435-1/c  
Sequence 1, Application PC/TUS9516435  
GENERAL INFORMATION:  
APPLICANT: The Johns Hopkins University School of Medicine  
TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/16435  
FILING DATE: 15-DEC-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/033W01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3807 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
PCT-US95-16435-1

Query Match 1.5%; Score 16; DB 5; Length 3807;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 gcgctggccttggcag 126

Db 352 GCGCTGGCCTTGGCAG 337  
|||||

Search completed: November 22, 2001, 01:52:17  
Job time: 5406 sec

---

This Page Blank (uspto)

**This Page Blank (usptc)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2001, 01:05:36 ; Search time 100.33 Seconds  
(without alignments)  
6649.059 Million cell updates/sec

Title: US-09-854-844-1  
Perfect score: 1041  
Sequence: 1 atgggcctgtgctgtgc.....ttattacaattgaaatga 1041

Scoring table:  
OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 475007 seqs, 224034647 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 950014

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Pending\_Patents\_NA\_New:\*  
1: /cgn2\_6/pdata/2/pna/PCT\_NEW\_COMB.seq:\*  
2: /cgn2\_6/pdata/2/pna/US06\_NEW\_COMB.seq:\*  
3: /cgn2\_6/pdata/2/pna/US07\_NEW\_COMB.seq:\*  
4: /cgn2\_6/pdata/2/pna/US08\_NEW\_COMB.seq:\*  
5: /cgn2\_6/pdata/2/pna/US09\_NEW\_COMB.seq:\*  
6: /cgn2\_6/pdata/2/pna/US60\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	1.7	434	5	US-09-922-340-5927
2	18	1.7	1413	5	US-09-805-467A-1
3	18	1.7	2079	5	US-09-525-993A-21
4	18	1.7	2300	5	US-09-805-467A-3
5	17	1.6	274	5	US-09-750-456-575
6	17	1.6	300	5	US-09-954-456-748
7	17	1.6	364	5	US-09-985-678-244637
8	17	1.6	376	5	US-09-985-678-98359
9	17	1.6	391	5	US-09-985-678-216751
10	17	1.6	445	5	US-09-969-347-298
11	17	1.6	850	5	US-09-954-456-748
12	17	1.6	6160	5	US-09-984-429-384
13	16	1.5	107	5	US-09-982-402-1865
14	16	1.5	228	5	US-09-985-678-510
15	16	1.5	229	5	US-09-985-678-116751
16	16	1.5	264	5	US-09-985-678-33227
17	16	1.5	268	5	US-09-985-678-192430
18	16	1.5	288	5	US-09-985-678-278670
19	16	1.5	290	5	US-09-985-678-168786
20	16	1.5	336	5	US-09-985-678-86530
21	16	1.5	376	5	US-09-985-678-236640
22	16	1.5	385	5	US-09-388-906A-21836
23	16	1.5	407	5	US-09-985-678-248101
24	16	1.5	413	5	US-09-922-340-1444
25	16	1.5	422	5	US-09-922-340-8279
26	16	1.5	425	5	US-09-985-678-269676
27	16	1.5	432	5	US-09-954-456-414

c 28 16 1.5 432 5 US-09-954-456-1411 Sequence 1411, Ap  
c 29 16 1.5 442 5 US-09-922-340-141 Sequence 141, App  
c 30 16 1.5 445 5 US-09-985-678-270107 Sequence 270107,  
c 31 16 1.5 447 5 US-09-985-678-99463 Sequence 99463, A  
c 32 16 1.5 448 5 US-09-985-678-92905 Sequence 92905, A  
c 33 16 1.5 449 5 US-09-985-678-247858 Sequence 247858,  
c 34 16 1.5 451 5 US-09-985-678-214012 Sequence 214012,  
c 35 16 1.5 452 5 US-09-922-340-3395 Sequence 3395, Ap  
c 36 16 1.5 460 5 US-09-388-906A-13021 Sequence 13021, A  
c 37 16 1.5 527 5 US-09-985-678-229097 Sequence 229097,  
c 38 16 1.5 534 5 US-09-985-678-222452 Sequence 222452,  
c 39 16 1.5 545 5 US-09-985-678-219737 Sequence 219737,  
c 40 16 1.5 551 5 US-09-985-678-228527 Sequence 228527,  
c 41 16 1.5 561 5 US-09-922-340-140 Sequence 140, App  
c 42 16 1.5 910 5 US-09-976-594-1068 Sequence 1068, Ap  
c 43 16 1.5 910 5 US-09-981-353-167 Sequence 167, App  
c 44 16 1.5 1534 5 US-09-620-394B-122 Sequence 122, App  
c 45 16 1.5 2016 5 US-09-896-594-15 Sequence 15, App

## ALIGNMENTS

RESULT 1  
US-09-922-340-5927  
; Sequence 5927, Application US/09922340  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained  
; FILE REFERENCE: 20411-726CON2  
; CURRENT APPLICATION NUMBER: US/09/922,340  
; CURRENT FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: 09/353,690  
; PRIOR FILING DATE: 1999-07-14  
; PRIOR APPLICATION NUMBER: US 09/004,182  
; PRIOR FILING DATE: 1998-01-07  
; NUMBER OF SEQ ID NOS: 12181  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5927  
; LENGTH: 434  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(434)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-922-340-5927

Query Match 1.7%; Score 18; DB 5; Length 434;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 ccaggatgctgtgcagg 110  
|||||  
Db 80 ccaggatgctgtgcagg 97

RESULT 2  
US-09-805-467A-1  
; Sequence 1, Application US/09805467A  
; GENERAL INFORMATION:  
; APPLICANT: Ramakrishnan, Shyam  
; TITLE OF INVENTION: Regulation of Human Lipoxin A4  
; FILE REFERENCE: 4974.00453  
; CURRENT APPLICATION NUMBER: US/09/805,467A  
; CURRENT FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 60/189,037  
; PRIOR FILING DATE: 2000-03-14  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0



Query Match 1.6%; Score 17; DB 5; Length 300;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 cctggaggtacaccaa 761  
|||||  
DB 281 cctggaggtacaccaa 297

## RESULT 7

US-09-985-678-244637  
; Sequence 244637, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US/09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 244637  
; LENGTH: 364  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-985-678-244637

Query Match 1.6%; Score 17; DB 5; Length 364;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 730 tctggtaaatctctcc 746  
|||||  
DB 306 tctggtaaatctctcc 322

## RESULT 8

US-09-985-678-98359  
; Sequence 98359, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US/09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 98359  
; LENGTH: 376  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-09-985-678-98359

Query Match 1.6%; Score 17; DB 5; Length 376;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 gctgctctctctgctgg 43  
|||||  
DB 348 gctgctctctctgctgg 364

## RESULT 9

US-09-985-678-216751/c  
; Sequence 216751, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; APPLICANT: Liu, Jingdong

; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US/09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 216751  
; LENGTH: 391  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-985-678-216751

Query Match 1.6%; Score 17; DB 5; Length 391;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 429 aattccaccctttgtt 445  
|||||  
DB 114 AATTCACCCCTTTGTT 98

## RESULT 10

US-09-969-347-298  
; Sequence 298, Application US/09969347  
; GENERAL INFORMATION:  
; APPLICANT: Ebner, Reinhard  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign  
; TITLE OF INVENTION: Sets  
; FILE REFERENCE: 689290-69  
; CURRENT APPLICATION NUMBER: US/09/969,347  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US/60/237,598  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: US/60/237,604  
; PRIOR FILING DATE: 2000-10-03  
; NUMBER OF SEQ ID NOS: 318  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 298  
; LENGTH: 445  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-969-347-298

Query Match 1.6%; Score 17; DB 5; Length 445;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gctggctgtgcttccac 26  
|||||  
DB 316 gctggctgtgcttccac 332

## RESULT 11

US-09-954-456-748  
; Sequence 748, Application US/09954456  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C  
; TITLE OF INVENTION: Sets  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25

```
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 748
; LENGTH: 850
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-748
```

```
Query Match 1.6%; Score 17; DB 5; Length 850;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 745 cctggagctctacacaa 761
Db 716 cctggagctctacacaa 732
|||||
```

```
RESULT 12
US-09-984-429-384/c
; Sequence 384, Application US/09984429
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P2
; CURRENT APPLICATION NUMBER: US/09/984,429
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,591
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/288,143
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: PCT/US98/21142
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/061,463
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,529
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/071,498
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,527
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,536
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,532
; PRIOR FILING DATE: 1997-10-09
; NUMBER OF SEQ ID NOS: 727
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 384
; LENGTH: 6160
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-984-429-384
```

```
Query Match 1.6%; Score 17; DB 5; Length 6160;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1025 ttacaattgaaatga 1041
Db 2046 TTTACAATTGAAATGA 2030
|||||
```

```
RESULT 13
US-09-982-402-1865
; Sequence 1865, Application US/09982402
; GENERAL INFORMATION:
; APPLICANT: [list inventors here]
; TITLE OF INVENTION: [list title here]
; FILE REFERENCE: [list docket ID here]
; CURRENT APPLICATION NUMBER: US/09/982,402
; CURRENT FILING DATE: 2001-10-17
; NUMBER OF SEQ ID NOS: 3553
; SOFTWARE: PERL Program
; SEQ ID NO 1865
; LENGTH: 107
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 700800104H1
; NAME/KEY: unsure
; LOCATION: 13
; OTHER INFORMATION: a, t, c, g, or other
US-09-982-402-1865
```

```
Query Match 1.5%; Score 16; DB 5; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 808 aacaatctagacttct 823
Db 35 aacaatctagacttct 50
|||||
```

```
RESULT 14
US-09-985-678-510/c
; Sequence 510, Application US/09985678
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 16517.255/38-21(15097)F
; CURRENT APPLICATION NUMBER: US/09/985,678
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 09/304,517
; PRIOR FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 295529
; SEQ ID NO 510
; LENGTH: 228
; TYPE: DNA
; ORGANISM: Zea mays
US-09-985-678-510
```

```
Query Match 1.5%; Score 16; DB 5; Length 228;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 918 agctgaagctgttgc 933
Db 88 AGCTGAAGCTGTGTGCT 73
|||||
```

```
RESULT 15
US-09-985-678-116751/c
; Sequence 116751, Application US/09985678
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 16517.255/38-21(15097)F
; CURRENT APPLICATION NUMBER: US/09/985,678
```

; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 116751  
; LENGTH: 229  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-985-678-116751

Query Match 1.5%; Score 16; DB 5; Length 229;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 ttatctatggaggttc 167  
|||||  
Db 86 TTATCTATGAGGTTTC 71

RESULT 16  
US-09-985-678-33227  
; Sequence 33227, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 33227  
; LENGTH: 264  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-09-985-678-33227

Query Match 1.5%; Score 16; DB 5; Length 264;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gctggctgtgccttca 25  
|||||  
Db 135 gctggctgtgccttca 150

RESULT 17  
US-09-985-678-192430  
; Sequence 192430, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 192430  
; LENGTH: 268  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-985-678-192430

Query Match 1.5%; Score 16; DB 5; Length 268;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 attctggaggcctct 671  
|||||  
Db 140 attctggaggcctct 155

RESULT 18  
US-09-985-678-278670  
; Sequence 278670, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; APPLICANT: Liu, Jingdong  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 278670  
; LENGTH: 288  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-985-678-278670

Query Match 1.5%; Score 16; DB 5; Length 288;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 ttatctatggaggttc 167  
|||||  
Db 140 ttatctatggaggttc 155

RESULT 19  
US-09-985-678-168786  
; Sequence 168786, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; APPLICANT: Liu, Jingdong  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 168786  
; LENGTH: 290  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-985-678-168786

Query Match 1.5%; Score 16; DB 5; Length 290;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 ttatctatggaggttc 167  
|||||  
Db 233 ttatctatggaggttc 248

RESULT 20  
US-09-985-678-86530  
; Sequence 86530, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; APPLICANT: Liu, Jingdong  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05

; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 86530  
; LENGTH: 336  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-09-985-678-86530

Query Match 1.5%; Score 16; DB 5; Length 336;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 918 agctgaagctgttgc 933  
|||||

Db 74 agctgaagctgttgc 89

## RESULT 21

US-09-985-678-236640  
; Sequence 236640, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; APPLICANT: Liu, Jingdong  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 236640  
; LENGTH: 376  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(376)  
; OTHER INFORMATION: unsure at all n locations  
US-09-985-678-236640

Query Match 1.5%; Score 16; DB 5; Length 376;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 ttatctatggaggttc 167  
|||||

Db 309 ttatctatggaggttc 324

## RESULT 22

US-09-388-906A-21836  
; Sequence 21836, Application US/09388906A  
; GENERAL INFORMATION:  
; APPLICANT: Havukkala, Ilkka  
; APPLICANT: Shenk, Michael Andrew  
; TITLE OF INVENTION: Polynucleotides Isolated From Plants and  
; TITLE OF INVENTION: Methods For Their Use  
; FILE REFERENCE: 11000.10130  
; CURRENT APPLICATION NUMBER: US/09/388,906A  
; CURRENT FILING DATE: 1999-09-01  
; NUMBER OF SEQ ID NOS: 24843  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21836  
; LENGTH: 385  
; TYPE: DNA  
; ORGANISM: Eucalyptus grandis  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(385)  
; OTHER INFORMATION: n = A,T,C or G

Query Match 1.5%; Score 16; DB 5; Length 413;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 ctgctctcttgcgtgg 43

## US-09-388-906A-21836

Query Match 1.5%; Score 16; DB 5; Length 385;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 aaaagttaaggaaagt 477  
|||||

Db 65 aaaagttaaggaaagt 80

## RESULT 23

US-09-985-678-248101  
; Sequence 248101, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; APPLICANT: Liu, Jingdong  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 248101  
; LENGTH: 407  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-985-678-248101

Query Match 1.5%; Score 16; DB 5; Length 407;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 851 tctctctgctctctc 866  
|||||

Db 14 tctctctgctctctc 29

## RESULT 24

US-09-922-340-1444/c  
; Sequence 1444, Application US/09922340  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained  
; FILE REFERENCE: 20411-726CON2  
; CURRENT APPLICATION NUMBER: US/09/922,340  
; CURRENT FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: 09/353,690  
; PRIOR FILING DATE: 1999-07-14  
; PRIOR APPLICATION NUMBER: US 09/004,182  
; PRIOR FILING DATE: 1998-01-07  
; NUMBER OF SEQ ID NOS: 12181  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1444  
; LENGTH: 413  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(413)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-922-340-1444

Query Match 1.5%; Score 16; DB 5; Length 413;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 ctgctctcttgcgtgg 43

Db 95 CTGCTCTTCTGCTGG 80  
|||||

## RESULT 25

US-09-922-340-8279/c  
; Sequence 8279, Application US/09922340  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained  
; FROM cDNA Libraries Of Fetal Liver-Spleen and Infant Brain  
; FILE REFERENCE: 20411-726CON2  
; CURRENT APPLICATION NUMBER: US/09/922,340  
; CURRENT FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: 09/353,690  
; PRIOR FILING DATE: 1999-07-14  
; PRIOR APPLICATION NUMBER: US 09/004,182  
; PRIOR FILING DATE: 1998-01-07  
; NUMBER OF SEQ ID NOS: 12181  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 8279  
; LENGTH: 422  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(422)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-922-340-8279

Query Match 1.5%; Score 16; DB 5; Length 422;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 ctgtctcttctgtctgg 43  
|||||

Db 289 CTGCTCTTCTGCTGG 274

## RESULT 26

US-09-985-678-269676  
; Sequence 269676, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; APPLICANT: Liu, Jingdong  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 269676  
; LENGTH: 425  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-985-678-269676

Query Match 1.5%; Score 16; DB 5; Length 425;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 789 tgcactatttcaaga 804  
|||||

Db 277 tgcactatttcaaga 292

## RESULT 27

US-09-954-456-414/c  
; Sequence 414, Application US/09954456  
; GENERAL INFORMATION:

; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 414  
; LENGTH: 432  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-456-414

Query Match 1.5%; Score 16; DB 5; Length 432;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 981 cagagaactcacagga 996  
|||||

Db 274 CAGAGAactCACAGGA 259

## RESULT 28

US-09-954-456-1411/c  
; Sequence 1411, Application US/09954456  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863

;  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1411  
; LENGTH: 432  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-456-1411

Query Match 1.5%; Score 16; DB 5; Length 432;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 981 cagagaactcacagga 996  
Db 274 CAGAGAACTCACAGGA 259  
|||||

RESULT 29  
US-09-922-340-141/c  
; Sequence 141, Application US/09922340  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained  
; FILE REFERENCE: From cDNA Libraries Of Fetal Liver-Spleen and Infant Brain  
; CURRENT APPLICATION NUMBER: US/09/922,340  
; PRIOR FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: 09/353,690  
; PRIOR FILING DATE: 1999-07-14  
; PRIOR APPLICATION NUMBER: US 09/004,182  
; PRIOR FILING DATE: 1998-01-07  
; NUMBER OF SEQ ID NOS: 12181  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 141  
; LENGTH: 442  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(442)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-922-340-141

Query Match 1.5%; Score 16; DB 5; Length 442;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 cgctgctctcttctgct 41  
Db 30 CGCTGCTCTCTTGTGCT 15  
|||||

RESULT 30  
US-09-985-678-270107  
; Sequence 270107, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 270107  
; LENGTH: 445  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-985-678-270107

Query Match 1.5%; Score 16; DB 5; Length 445;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 ttatctatggaggttc 167  
Db 323 ttatctatggaggttc 338  
|||||

RESULT 31  
US-09-985-678-99463/c  
; Sequence 99463, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 99463  
; LENGTH: 447  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-09-985-678-99463

Query Match 1.5%; Score 16; DB 5; Length 447;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 918 agctgaagctgttgct 933  
Db 63 AGCTGAAGCTGTGCT 48  
|||||

RESULT 32  
US-09-985-678-92905  
; Sequence 92905, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 92905  
; LENGTH: 448  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-09-985-678-92905

Query Match 1.5%; Score 16; DB 5; Length 448;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gctggctgtgccttca 25  
Db 69 gctggctgtgccttca 84  
|||||

RESULT 33  
US-09-985-678-247858  
; Sequence 247858, Application US/09985678  
; GENERAL INFORMATION:



```
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 16517.255/38-21(15097)F
; CURRENT APPLICATION NUMBER: US/09/985,678
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 09/304,517
; PRIOR FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 295529
; SEQ ID NO 247858
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(449)
; OTHER INFORMATION: unsure at all n locations
US-09-985-678-247858
```

```
Query Match 1.5%; Score 16; DB 5; Length 449;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 916 gtactggaagctgttg 931
|||||
DB 40 gtactggaagctgttg 55
```

```
RESULT 34
US-09-985-678-214012/c
; Sequence 214012, Application US/09985678
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 16517.255/38-21(15097)F
; CURRENT APPLICATION NUMBER: US/09/985,678
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 09/304,517
; PRIOR FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 295529
; SEQ ID NO 214012
; LENGTH: 451
; TYPE: DNA
; ORGANISM: Glycine max
US-09-985-678-214012
```

```
Query Match 1.5%; Score 16; DB 5; Length 451;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 152 ttatctatggaggttc 167
|||||
DB 284 TTATCTATGAGGTTTC 269
```

```
RESULT 35
US-09-922-340-3395/c
; Sequence 3395, Application US/09922340
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
; FILE REFERENCE: 20411-726CON2
; CURRENT APPLICATION NUMBER: US/09/922,340
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/353,690
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 09/004,182
; PRIOR FILING DATE: 1998-01-07
; NUMBER OF SEQ ID NOS: 12181
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3395
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(452)
; OTHER INFORMATION: n = A,T,C or G
US-09-922-340-3395
```

```
Query Match 1.5%; Score 16; DB 5; Length 452;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 494 attctgcccttcagga 509
|||||
DB 214 ATTCTGCCCTTCAGGA 199
```

```
RESULT 36
US-09-388-906A-19021
; Sequence 19021, Application US/09388906A
; GENERAL INFORMATION:
; APPLICANT: Havukkala, Ilkka
; APPLICANT: Shenk, Michael Andrew
; TITLE OF INVENTION: Polynucleotides Isolated From Plants and
; FILE REFERENCE: 11000.1013U
; CURRENT APPLICATION NUMBER: US/09/388,906A
; CURRENT FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 24843
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19021
; LENGTH: 460
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-388-906A-19021
```

```
Query Match 1.5%; Score 16; DB 5; Length 460;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 462 aaaagtttaaggaagt 477
|||||
DB 53 aaaagtttaaggaagt 68
```

```
RESULT 37
US-09-985-678-229097/c
; Sequence 229097, Application US/09985678
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 16517.255/38-21(15097)F
; CURRENT APPLICATION NUMBER: US/09/985,678
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 09/304,517
; PRIOR FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 295529
; SEQ ID NO 229097
; LENGTH: 527
; TYPE: DNA
; ORGANISM: Glycine max
US-09-985-678-229097
```

```
Query Match 1.5%; Score 16; DB 5; Length 527;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 152 ttatctatggaggctt 167  
|||||  
Db 277 TTATCTATGGAGGCTC 262

RESULT 38  
US-09-985-678-222452/c  
; Sequence 222452, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 222452  
; LENGTH: 534  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-985-678-222452

Query Match 1.5%; Score 16; DB 5; Length 534;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 ttatctatggaggctt 167  
|||||  
Db 439 TTATCTATGGAGGCTC 424

RESULT 39  
US-09-985-678-219737  
; Sequence 219737, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678  
; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 219737  
; LENGTH: 545  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-985-678-219737

Query Match 1.5%; Score 16; DB 5; Length 545;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 948 ggaagagaatgcattgg 963  
|||||  
Db 495 ggaagagaatgcattgg 510

RESULT 40  
US-09-985-678-228527/c  
; Sequence 228527, Application US/09985678  
; GENERAL INFORMATION:  
; APPLICANT: Cheikh, Nordine  
; TITLE OF INVENTION: Annotated Plant Genes  
; FILE REFERENCE: 16517.255/38-21(15097)F  
; CURRENT APPLICATION NUMBER: US/09/985,678

; CURRENT FILING DATE: 2001-11-05  
; PRIOR APPLICATION NUMBER: US 09/304,517  
; PRIOR FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 295529  
; SEQ ID NO 228527  
; LENGTH: 551  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-985-678-228527

Query Match 1.5%; Score 16; DB 5; Length 551;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 816 agactctctctgacttc 831  
|||||  
Db 46 AGACTTCTCTGACTTC 31

RESULT 41  
US-09-922-340-140/c  
; Sequence 140, Application US/09922340  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained  
; FILE REFERENCE: 20411-726CON2  
; CURRENT APPLICATION NUMBER: US/09/922,340  
; CURRENT FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: 09/353,690  
; PRIOR FILING DATE: 1999-07-14  
; PRIOR APPLICATION NUMBER: US 09/004,182  
; PRIOR FILING DATE: 1998-01-07  
; NUMBER OF SEQ ID NOS: 12181  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 140  
; LENGTH: 561  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(561)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-922-340-140

Query Match 1.5%; Score 16; DB 5; Length 561;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 cgctgctctcttctgct 41  
|||||  
Db 30 CGCTGCTCTCTCTGCT 15

RESULT 42  
US-09-976-594-1068  
; Sequence 1068, Application US/09976594  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROI  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 1068  
; LENGTH: 910  
; TYPE: DNA

; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No: 206344.1  
US-09-976-594-1068

Query Match 1.5%; Score 16; DB 5; Length 910;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 cacgctgctcctctg 39  
|||||  
Db 113 cacgctgctcctctg 128

RESULT 43  
US-09-981-353-167  
; Sequence 167, Application US/09981353  
; GENERAL INFORMATION:  
; APPLICANT: Lasek, Amy W.  
; APPLICANT: Jones, David A.  
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER  
; FILE REFERENCE: PA-0038 US  
; CURRENT APPLICATION NUMBER: US/09/981,353  
; CURRENT FILING DATE: 2001-10-11  
; NUMBER OF SEQ ID NOS: 194  
; SOFTWARE: PERL Program  
; SEQ ID NO 167  
; LENGTH: 910  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No: 206344.1  
US-09-981-353-167

Query Match 1.5%; Score 16; DB 5; Length 910;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 cacgctgctcctctg 39  
|||||  
Db 113 cacgctgctcctctg 128

RESULT 44  
US-09-620-394B-122/c  
; Sequence 122, Application US/09620394B  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nickolai  
; APPLICANT: BROVER, Vyacheslav  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE OF INVENTION: Thereby  
; FILE REFERENCE: 2750-1067P  
; CURRENT APPLICATION NUMBER: US/09/620,394B  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 9131  
; SEQ ID NO 122  
; LENGTH: 1534  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: 1..1534  
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other  
; NAME/KEY: misc.feature  
; LOCATION: 1..1534  
; OTHER INFORMATION: Ceres Seq. ID 1375365  
US-09-620-394B-122

Query Match 1.5%; Score 16; DB 5; Length 1534;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 599 aggaagacaagaatttg 614  
|||||  
Db 983 AGCAGACAAGATTTC 968

RESULT 45  
US-09-896-594-15/c  
; Sequence 15, Application US/09896594  
; GENERAL INFORMATION:  
; APPLICANT: Contreras, Roland  
; APPLICANT: Callewaert, Nico L. M.  
; APPLICANT: Geysens, Steven C. J.  
; TITLE OF INVENTION: PROTEIN GLYCOSYLATION MODIFICATION IN PICHIA PASTORIS  
; FILE REFERENCE: 13748  
; CURRENT APPLICATION NUMBER: US/09/896,594  
; CURRENT FILING DATE: 2001-06-29  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 2016  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: The ORF  
; OTHER INFORMATION: sequence of the MFmManHDEL fusion in  
; OTHER INFORMATION: pGAP2MFmManHDEL.  
US-09-896-594-15

Query Match 1.5%; Score 16; DB 5; Length 2016;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 gccaggatgctgctgc 107  
|||||  
Db 1079 GCCAGGATGCTGCTGC 1064

Search completed: November 22, 2001, 01:55:19  
Job time: 2983 sec

---

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2001, 00:19:46 ; Search time 1179.76 Seconds  
(without alignments)  
8341.036 Million cell updates/sec

Title: US-09-854-844-1  
Perfect score: 1041  
Sequence: 1 atgggcctctgctgtgc.....ttattacaattgaaatga 1041

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 0

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est33:\*  
26: gb\_est34:\*  
27: gb\_est35:\*  
28: gb\_est36:\*  
29: gb\_est37:\*  
30: gb\_est38:\*  
31: gb\_est39:\*  
32: gb\_est40:\*  
33: em\_estba:\*  
34: em\_estfun:\*  
35: em\_esthum1:\*  
36: em\_esthum2:\*  
37: em\_esthum3:\*  
38: em\_esthum4:\*  
39: em\_esthum5:\*  
40: em\_esthum6:\*  
41: em\_esthum7:\*  
42: em\_esthum8:\*  
43: em\_esthum9:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est41:\*  
111: gb\_est42:\*  
112: gb\_est43:\*  
113: gb\_est44:\*  
114: gb\_est45:\*  
115: gb\_est46:\*  
116: gb\_est47:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	22	2.1	858	221	CNS03G2C	AL242445 Tetraodon
2	21	2.0	425	224	AQ108514	CIT-HSP-2
3	21	2.0	461	233	AQ770483	HS_5372_A
4	21	2.0	495	148	BF391435	UI-R-CAL-
5	21	2.0	601	118	AW621676	EST312474
6	21	2.0	687	119	AW649926	EST328380
7	21	2.0	699	119	AW650706	EST329160
8	21	2.0	706	118	AW622168	EST312966
9	21	2.0	789	239	AZ211775	SP_0153_B
10	20	1.9	472	238	AZ077203	RPCI-23-4
11	20	1.9	521	225	AQ178417	HS_2223_A
12	20	1.9	525	220	CNS02BW7	AL190384 Tetraodon
13	20	1.9	543	241	AZ332113	IM0060L14
14	20	1.9	566	236	AQ987586	RPCI-23-2
15	20	1.9	912	221	CNS047EX	Tetraodon
16	20	1.9	927	220	CNS032DS	AL224713 Tetraodon
17	20	1.9	977	221	CNS041MQ	AL270395 Tetraodon
18	20	1.9	1043	220	CNS02HOP	AL197026 Tetraodon
19	19	1.8	199	28	AV280134	AV280134
20	19	1.8	234	246	AZ626302	IM0466B06
21	19	1.8	307	3	AA167980	ms29q03_r
22	19	1.8	309	147	BF332578	BF332578 PM4-BF072
23	19	1.8	320	20	A1450971	ms29d03.x
24	19	1.8	358	148	BF403284	UI-R-CAL-
25	19	1.8	359	228	AQ444260	GSTC0303
26	19	1.8	367	152	BG314065	WHE2467_F
27	19	1.8	372	150	BF523355	UI-R-GO-U
28	19	1.8	401	235	AQ090926	GSTC0987
29	19	1.8	417	225	AQ178480	HS_2222_B
30	19	1.8	418	22	A1562186	vw74e07.x
31	19	1.8	438	12	AA855540	vw74e07.r
32	19	1.8	463	238	TA62C12Q	BE133527
33	19	1.8	471	163	BE133527	ug34b11-y
34	19	1.8	506	228	AQ413533	RPCI-11-1
35	19	1.8	515	231	AQ646668	RPCI93-DP
36	19	1.8	523	169	BF756411	PM4-CF011
37	19	1.8	527	119	AW650509	EST328963
38	19	1.8	552	162	BE030483	128575 MA
39	19	1.8	563	23	AT644407	v192a05.x
40	19	1.8	590	226	AQ307958	CIT-HSP-2
41	19	1.8	596	225	AQ197598	CIT-HSP-2
42	19	1.8	621	235	AQ923548	RPCI-23-2
43	19	1.8	627	232	AQ272580	HS_5456_A
44	19	1.8	664	251	AZ913822	RPCI-24-1
45	19	1.8	697	145	BF167775	601773578

## ALIGNMENTS

RESULT 1	CNS03G2C	CNS03G2C	858 bp	DNA	GSS	17-MAY-2000
LOCUS	Tetraodon nigroviridis genome survey sequence T7 end of clone 023H05 of library G from Tetraodon nigroviridis, genomic survey sequence.					
DEFINITION	AL242445.1 GI:7963214					
ACCESSION	GSS: genome survey sequence.					
VERSION	Tetraodon nigroviridis					
KEYWORDS	Tetraodon nigroviridis					
SOURCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.					
ORGANISM	1 (bases 1 to 858)					
REFERENCE	Roest-Crolius, H., Jaillon, O., Dasilva, C., Fitzames, C., Fisher, C.,					
AUTHORS						

Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.  
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  
Unpublished  
2 (bases 1 to 858)  
Roest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fitzames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.  
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence  
Unpublished  
3 (bases 1 to 858)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

## FEATURES

source  
1..858  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="023H05"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COBG023CD03LP1-end : T7"

BASE COUNT 194 a 233 c 213 g 205 t 13 others  
ORIGIN

Query Match 2.1%; Score 22; DB 221; Length 858;  
Best Local Similarity 100.0%; Pred. No. 0.86; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 gctggccttggcaggtcagcct 134  
|||||

Db 736 GCTGGCCTTGGCAGTCAGCCT 757  
|||||

RESULT 2

AQ108514

LOCUS

DEFINITION CIT-HSP-2379E23.TF CIT-HSP Homo sapiens genomic clone 2379E23, DNA

sequence.

ACCESSION AQ108514

VERSION AQ108514.1

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 425)

Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.

Use of a random human BAC End Sequence Database for Sequence-Ready

Map Building

Unpublished (1998)

Other\_GSSs: CIT-HSP-2379E23.TF

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

[http://www.tigr.org/tldb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html).

Seq primer: M13-21

Class: BAC ends.

Location/Qualifiers

source

1. .425  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2379E23"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2:  
HindIII"

BASE COUNT 99 a 111 c 107 g 108 t

ORIGIN

Query Match 2.0%; Score 21; DB 224; Length 425;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 cactctgcctcctgcctat 398

|||||

Db 85 CACTTCTGCATCCTGCCTAT 105

RESULT 3

AQ770483/c

LOCUS  
DEFINITION HS\_5372\_A2\_F02.SP6E.RPCI-11 Human Male BAC Library Homo sapiens  
genomic clone Plate=948 Col=4 Row=K, DNA sequence.

ACCESSION AQ770483.1 GI:5648599

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 461)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieterdejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)

or from Resear h Genetics (info@resgen.com). BAC end Web Server:

http://www.htsc.washington.edu

Plate: 948 row: K column: 4

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 461.

Location/Qualifiers

1. .461

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate=948 Col=4 Row=K"

/clone\_lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;

Male blood DNA was isolated from one randomly chosen donor

and partially digested with a combination of EcoRI and

EcoRI Methylase. Size selected DNA was cloned into the

pBAC3.6 vector at EcoRI sites"

114 a 119 c 86 g 142 t

BASE COUNT

ORIGIN

Query Match 2.0%; Score 21; DB 233; Length 461;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 649 aagggtgattctggaggcct 669

|||||

Db 374 AAGGTTGATTCCTGGAGGCCT 354

RESULT 4

BF391435/c

LOCUS

DEFINITION UI-R-CAL-bcx-d-10-0-UI.s1 UI-R-CAL Rattus norvegicus cDNA clone  
UI-R-CAL-bcx-d-10-0-UI 3', mRNA sequence.

ACCESSION BF391435

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 495)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL 9704477

MEDLINE

COMMENT

Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to identify it as a clone from the  
normalized hypothalamus library cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com) The following repetitive  
elements were found in this cDNA sequence: 7-360, >RMER1B#Unknown  
Seq primer: M13 Forward  
POLYA=Yes.

Location/Qualifiers

1. .495

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-CAL-bcx-d-10-0-UI"

/clone\_lib="UI-R-CAL"

/lab\_host="DHI08 (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-CAL

library is a subtracted library derived from the following

tissues: thalamus, cerebellum, hypothalamus, medulla, pons

, midbrain, cerebral cortex, corpus striatum, testis, and

hippocampus. For a detailed description of the library

from which this clone was derived, please visit our web

site at ratest.eng.uiowa.edu. The subtraction has been

previously described in (Bonaldo, Lennon and Soares,

Genome Research 6:791-806, 1996)

TAG\_LIB=UI-R-CAL

TAG\_TISSUE=hypothalamus

TAG\_SEQ=GATGC"

105 a 135 c 95 g 160 t

BASE COUNT

ORIGIN



```

Query Match      2.0%; Score 21; DB 148; Length 495;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 402 ctggccaggtcacaaagca 422
|||||
Db 443 CTGGCCAGGTGCACAAAGCA 423

RESULT 5
AW621676      601 bp      mRNA      28-MAR-2000
LOCUS EST312474 tomato root during/after fruit set, Cornell University
DEFINITION Lycopersicon esculentum cDNA clone CLEX13A21 5', mRNA sequence.
ACCESSION AW621676
VERSION AW621676.1 GI:7333323
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 601)
AUTHORS van der Hoeven,R.S., Garvin,D., Matern,A.L., Holt,I.E., Kochian,L.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
,S.D.
Generation of ESTs from tomato root, during and after fruit set
Unpublished (1999)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
source 1..601
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEX13A21"
/clone_lib="tomato root during/after fruit set, Cornell
University"
/tissue_type="root"
/dev_stage="plants during and after fruit-set"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Tanksley; Tissue supplied by Dave Garvin
(USDA-ARS, Ithaca, NY 14850)."
```

---

```

BASE COUNT 173 a 97 c 108 g 223 t
ORIGIN

Query Match      2.0%; Score 21; DB 118; Length 601;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 781 tggattaatgccactatttca 801
|||||
Db 8 TGGATTAATGCCACTATTTC A 28

RESULT 6
AW649926      687 bp      mRNA      04-APR-2000
LOCUS EST328380 tomato germinating seedlings, TAMU Lycopersicon
DEFINITION esculentum cDNA clone CLEILK13 5', mRNA sequence.
ACCESSION AW649926
VERSION AW649926.1 GI:7411164
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 699)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley
,S.D.
Generation of ESTs from germinating tomato seed
Unpublished (2000)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
source 1..687
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEILK13"
/clone_lib="tomato germinating seedlings, TAMU"
/tissue_type="whole seedlings"
/dev_stage="7 days post imbibition"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; 7 days post imbibition on water-agar. Mixed stage
whole germinating seedlings from seed coat emergence up
to two centimeters in seeds not showing obvious signs of
germination were discarded."
```

---

```

BASE COUNT 223 a 119 c 119 g 226 t
ORIGIN

Query Match      2.0%; Score 21; DB 119; Length 687;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 781 tggattaatgccactatttca 801
|||||
Db 378 TGGATTAATGCCACTATTTC A 398

RESULT 7
AW650706      699 bp      mRNA      04-APR-2000
LOCUS EST329160 tomato germinating seedlings, TAMU Lycopersicon
DEFINITION esculentum cDNA clone CLEILJ38 5', mRNA sequence.
ACCESSION AW650706
VERSION AW650706.1 GI:7411944
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 699)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley
,S.D.
Generation of ESTs from germinating tomato seed
Unpublished (2000)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
source 1..687
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEILJ38"
/clone_lib="tomato germinating seedlings, TAMU"
/tissue_type="whole seedlings"
/dev_stage="7 days post imbibition"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; 7 days post imbibition on water-agar. Mixed stage
whole germinating seedlings from seed coat emergence up
to two centimeters in seeds not showing obvious signs of
germination were discarded."
```

Fax: 864 656 4293  
Email: dfrisch@CLEMSON.EDU  
5 prime sequence.

**FEATURES**  
source  
Location/Qualifiers  
1..699  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLE11378"  
/clone\_lib="tomato germinating seedlings, TAMU"  
/issue\_type="whole seedlings"  
/dev\_stage="7 days post imbibition"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; 7 days post imbibition on water-agar. Mixed stage whole germinating seedlings from seed coat emergence up to two centimeters in seeds not showing obvious signs of germination were discarded."  
232 a 125 c 116 g 225 t 1 others

**BASE COUNT**  
232 a 125 c 116 g 225 t

**ORIGIN**  
Query Match 2.0%; Score 21; DB 119; Length 699;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 781 tggattaatgccactatttca 801  
|||||  
Db 422 TGGATTAAATGCCACTATTTC 442

**RESULT 8**  
AW622168 706 bp mRNA EST 28-MAR-2000  
LOCUS  
DEFINITION  
Lycopersicon esculentum cDNA clone cLEX1402 5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.  
REFERENCE  
1 (bases 1 to 706)  
van der Hoeven,R.S., Garvin,D., Matern,A.L., Holt,I.E., Kochian,L., Liang,F., Hansen,T., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.  
Generation of ESTs from tomato root, during and after fruit set  
Unpublished (1999)  
Contact: David Frisch  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 4366  
Fax: 864 656 4293  
Email: dfrisch@CLEMSON.EDU  
5 prime sequence.

**FEATURES**  
source  
Location/Qualifiers  
1..706  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEX1402"  
/clone\_lib="tomato root during/after fruit set, Cornell University"  
/issue\_type="root"  
/dev\_stage="plants during and after fruit-set"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; supplier: Tanksley; Tissue supplied by Dave Garvin (USDA-ARS, Ithaca, NY 14850)."  
255 a 108 c 110 g 232 t 1 others

**BASE COUNT**  
255 a 108 c 110 g 232 t

**ORIGIN**  
Query Match 2.0%; Score 21; DB 118; Length 706;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 781 tggattaatgccactatttca 801  
|||||  
Db 683 TGGATTAAATGCCACTATTTC 703

**RESULT 9**  
AZ211775 789 bp DNA GSS 31-AUG-2000  
LOCUS  
DEFINITION  
SP\_0153\_B2.C05.T7A Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate-153 Col-10 Row=F, DNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Strongylocentrotus purpuratus.  
Strongylocentrotus purpuratus  
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinozoa; Echinoidae; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.  
REFERENCE  
1 (bases 1 to 789)  
Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Blondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.  
A sea urchin genome project: Sequence scan, virtual map, and additional resources  
Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)  
20402566  
Contact: Cameron, RA, Davidson, EH, Hood, L  
Division of Biology 156-29  
California Institute of Technology  
Pasadena California 91125, USA  
Tel: (626) 395-8421  
Fax: (626) 793-3047  
Email: acameron@caltech.edu  
Plate: 153 row: F column: 10  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 789.  
Location/Qualifiers  
1..789  
/organism="Strongylocentrotus purpuratus"  
/db\_xref="taxon:7668"  
/clone="Plate=153 Col=10 Row=F"  
/clone\_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"  
/note="organ: sperm; Vector: BACE3.6; BAC Clones in E-Coli DH10B"  
200 a 164 c 184 g 241 t

**BASE COUNT**  
200 a 164 c 184 g 241 t

**ORIGIN**  
Query Match 2.0%; Score 21; DB 239; Length 789;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 493 cattctgcccttcaggaagca 513  
|||||  
Db 471 CATTCTGCCCTTCAGGAAGCA 451

**RESULT 10**  
AZ077203 472 bp DNA GSS 31-MAR-2000  
LOCUS  
DEFINITION  
RPCI-23-407J4.TJ RPCI-23 Mus musculus genomic clone RPCI-23-407J4, DNA sequence.

```

ACCESSION AZ077203
VERSION AZ077203.1 GI:7370192
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 472)
AUTHORS Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Zhao, S., Nierman, M., Feldblyum, T., Malek, J., Shatsman, S., Akinret
, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 407 row: J column: 4
Seq primer: SP6
Class: BAC ends.
FEATURES             source
    Location/Qualifiers
        1..472
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="RPCI-23-407J4"
            /clone_lib="RPCI-23"
            /sex="Female"
            /lab_host="DH10B"
            /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site:1:
            EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or
            brain genomic DNA was isolated and partially digested
            with a combination of EcoRI and EcoRI Methylase. Size
            selected DNA was cloned into the pBACE3.6 vector at the
            EcoRI sites. The ligation products were transformed into
            DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      88 a 122 c 133 g 129 t
ORIGIN

Query Match      1.9%; Score 20; DB 238; Length 472;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 655 gattctggaggccctctgtc 674
|||||
DB 103 GATTCTGGAGGCGCTGTGTC 122

RESULT 11
LOCUS AQ178417
DEFINITION HS_2223_AL_D01_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-2223 Col-1 Row-G, DNA sequence.
ACCESSION AQ178417
VERSION AQ178417.1 GI:3575784
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 521)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2223 row: G column: 1
Class: BAC ends
High quality sequence stop: 521.
FEATURES             source
    Location/Qualifiers
        1..521
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Plate-2223 Col-1 Row-G"
            /clone_lib="CIT Approved Human Genomic Sperm Library D"
            /sex="male"
            /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
            E-Coli DH10B"
BASE COUNT      131 a 88 c 146 g 135 t 21 others
ORIGIN

Query Match      1.9%; Score 20; DB 225; Length 521;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 aatggattaatgccactatt 798
|||||
DB 170 AATGGATTAAATGCCACTATT 189

RESULT 12
LOCUS CNS02BW7
DEFINITION Tetraodon nigroviridis genome survey sequence pUC-Ori end of clone
254C18 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL190384
VERSION AL190384.1 GI:7828488
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 525)
AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bonneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 525)
AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Bonneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 525)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis

```

genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.

## FEATURES

source  
1. .525  
Location/Qualifiers  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG254BB09SP1-end :  
PUC-ori"

BASE COUNT 75 a 176 c 164 g 105 t 5 others  
ORIGIN

Query Match 1.9%; Score 20; DB 220; Length 525;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 tggcctggcaggtcagcct 134

Db 273 TGGCTTGGCAGGTACGCT 292

## RESULT 13

AZ332113 543 bp DNA GSS 29-SEP-2000  
LOCUS  
DEFINITION IM0060L14F Mouse 10kb plasmid UUGCIM library Mus musculus genomic  
clone UUGCIM0060L14 F, DNA sequence.

ACCESSION AZ332113

VERSION AZ332113.1 GI:10395455

KEYWORDS GSS.

SOURCE house mouse.

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 543)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0060 row: L column: 14

Seq primer: CGTGTAAACGACGCGCAGT

Class: plasmid ends

High quality sequence stop: 543.

## FEATURES

source

1. .543  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGCIM0060L14"  
/clone\_lib="Mouse 10kb plasmid UUGCIM library"  
/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (g114732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 163 a 106 c 133 g 141 t

## ORIGIN

Query Match 1.9%; Score 20; DB 241; Length 543;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 tgtcacaaagcagttggcaa 430

Db 361 TGTCAACAAGCAGTTGGCAA 380

## RESULT 14

AQ987586 566 bp DNA GSS 30-JAN-2000  
LOCUS  
DEFINITION RPCI-23-282C1.TV RPCI-23 Mus musculus genomic clone RPCI-23-282C1,  
DNA sequence.

ACCESSION AQ987586

VERSION AQ987586.1 GI:6820791

KEYWORDS GSS.

SOURCE house mouse.

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 566)  
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret,  
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Kroi, M., de Jong, P.  
and Fraser, C.M.

Mouse BAC end Sequences from Library RPCI-23

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)  
or from Resea ch Genetics (info@resgen.com). BAC end page:  
http://www.tigr.org/tdb/bacends/mouse/bac\_end\_intro.html

Plate: 282 row: C column: 1

Seq primer: T7

Class: BAC ends.

## FEATURES

source

Location/Qualifiers  
1. .566  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-282C1"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1:  
ECORI; Site\_2: ECORI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of ECORI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACE3.6 vector at the  
ECORI sites. The ligation products were transformed into

```
CNS032DS      927 bp      DNA      GSS      15-MAY-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
206824 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION      AL224713
VERSION        GI:7883586
KEYWORDS       GSS: genome survey sequence.
SOURCE         Tetraodon nigroviridis.
ORGANISM       Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE      1 (bases 1 to 927)
AUTHORS        Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE          Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL        Unpublished
AUTHORS        2 (bases 1 to 927)
REFERENCE      Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bertot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE          Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL        Unpublished
AUTHORS        3 (bases 1 to 927)
REFERENCE      Genoscope.
TITLE          Direct Submission
JOURNAL        Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT        This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
Source
1..927
/organism="Tetraodon nigroviridis"
/db_xref="taxon:9883"
/clone="206824"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG206DAL2LP1-end : T7"

BASE COUNT    183 a 272 c 274 g 197 t 1 others
ORIGIN

Query Match    1.9%; Score 20; DB 220; Length 927;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 tggccttggcaggtcacgct 134
|||||
Db 272 TGGCCTTGGCAGGTCAAGCCT 291

RESULT 17
CNS041MQ      977 bp      DNA      GSS      18-MAY-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
074B16 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION      AL270395
VERSION        GI:7992313
KEYWORDS       GSS: genome survey sequence.
SOURCE         Tetraodon nigroviridis.
ORGANISM       Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE      1 (bases 1 to 977)
AUTHORS        Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
```

```
CNS047EX      912 bp      DNA      GSS      18-MAY-2000
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
087M02 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION      AL277890
VERSION        GI:8012093
KEYWORDS       GSS: genome survey sequence.
SOURCE         Tetraodon nigroviridis.
ORGANISM       Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE      1 (bases 1 to 912)
AUTHORS        Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE          Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL        Unpublished
AUTHORS        2 (bases 1 to 912)
REFERENCE      Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bertot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE          Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL        Unpublished
AUTHORS        3 (bases 1 to 912)
REFERENCE      Genoscope.
TITLE          Direct Submission
JOURNAL        Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT        This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
Location/Qualifiers
1..912
/organism="Tetraodon nigroviridis"
/db_xref="taxon:9883"
/clone="087M02"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG087BG01SP1-end :
PUC-Ori"

BASE COUNT    173 a 251 c 260 g 222 t 6 others
ORIGIN

Query Match    1.9%; Score 20; DB 221; Length 912;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 tggccttggcaggtcacgct 134
|||||
Db 833 TGGCCTTGGCAGGTCAAGCCT 852

RESULT 16
```



19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES  
source  
1. .199  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="4933412119"  
/clone\_lib="RIKEN full-length enriched, adult male testis (DH10B)"  
/sex="male"  
/tissue\_type="testis"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
GAGAGAGAGATCCAGACGCTCTTTTTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'  
GAGAGAGATCTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC 1. Cloning sites, 5' end: SalI; 3' end: BamHI."

BASE COUNT 28 a 69 c 45 g 57 t  
ORIGIN

Query Match 1.8%; Score 19; DB 28; Length 199;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 ttgggtgaccggatggga 462  
|||||  
Db 58 TTGGGTGACCGGATGGGA 40

RESULT 20  
AZ626302  
LOCUS  
DEFINITION  
IM0466B06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0466B06 R, DNA sequence.  
AZ626302  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 234)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.  
TITLE  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL  
Unpublished (2000)  
COMMENT  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
Insert Length: 10000 Std Error: 0.00

Plate: 0466 row: B column: 06  
Seq primer: CACACAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 234.  
Location/Qualifiers  
1. .234  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0466B06"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 46 a 60 c 57 g 71 t  
ORIGIN

Query Match 1.8%; Score 19; DB 246; Length 234;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 667 cctctgtcgtgtcacattg 685  
|||||  
Db 15 CCTCTGCTGTCACATTG 33

RESULT 21  
AA167980/C  
LOCUS  
DEFINITION  
ms29d03.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone IMAGE:608357 5', mRNA sequence.  
AA167980  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 307)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.  
TITLE  
The WashU-HMI Mouse EST Project  
JOURNAL  
Unpublished (1996)  
COMMENT  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [mousseest@watson.wustl.edu](mailto:mousseest@watson.wustl.edu)  
This clone is available royalty-free through LML; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.  
 MG1:373789

Seq primer: -28ml3 rev1 ET from Amersham.

FEATURES  
 source

Location/Qualifiers  
 1. .307  
 /organism="Mus musculus"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone\_lib="Stratagene mouse skin (#937313)"  
 /sex="females"  
 /tissue\_type="whole skin"  
 /dev\_stage="11 weeks old"

/lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: skin; Vector: pBluescript SK-; Site\_1: EcoRI  
 ; Site\_2: XhoI; Cloned unidirectionally, primer: Oligo  
 dr. Whole skin from 11 week old C57BL/6 female mice.  
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
 adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor  
 sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3"

BASE COUNT 60 a 81 c 102 g 64 t  
 ORIGIN

Query Match 1.8%; Score 19; DB 3; Length 307;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 88 ggtgcccagatgctgctg 106  
 |||||  
 Db 91 GGTGCCAGGATGCTGCTG 73

RESULT 22  
 BF332578/c

LOCUS 309 bp mRNA EST 22-NOV-2000  
 DEFINITION PM4-BT0724-010400-008-c10 BT0724 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BF332578  
 VERSION BF332578.1 GI:11303326  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 309)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&st2=PM4-BT0724-  
 010400-008-cl0&st3=2000-04-01&st4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 18  
 High quality sequence stop: 32.  
 Location/Qualifiers  
 1. .309  
 /organism="Homo sapiens"

FEATURES  
 source

Location/Qualifiers  
 1. .309  
 /organism="Homo sapiens"

/db\_xref="taxon:9606"  
 /clone\_lib="BT0724"

/dev\_stage="Adult"  
 /note="Organ: breast; Vector: puc18; Site\_1: SmaI; Site\_2:  
 SmaI; A mini-library was made by cloning products derived  
 from ORESTES PCR (U.S. Letters Patent application No. 196  
 ,716 - Ludwig Institute for Cancer Research) profiles  
 into the puc 18 vector. Reverse transcription of tissue  
 mRNA and cDNA amplification were performed under low  
 stringency conditions."

BASE COUNT 95 a 64 c 73 g 77 t  
 ORIGIN

Query Match 1.8%; Score 19; DB 147; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 623 atactcaaaacatgaagga 641  
 |||||  
 Db 283 ATACTCAAAACATGAAGGA 265

RESULT 23  
 AI450971

LOCUS 320 bp mRNA EST 09-MAR-1999  
 DEFINITION ms29d03.x1 Stratagene mouse skin (#937313) Mus musculus cDNA clone  
 IMAGE:608357 3' similar to SW:RGSC\_RAT 008774 REGULATOR OF  
 G-PROTEIN SIGNALLING 12 ;, mRNA sequence.

ACCESSION AI450971  
 VERSION AI450971.1 GI:4300731  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 320)  
 Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person,  
 B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter,  
 E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
 Waterston, R., and Wilson, R.  
 The WashU-NCI Mouse EST Project 1999  
 Unpublished (1999)  
 Contact: Marra M/WashU-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 This clone was previously sequenced on the 5' end only, this new  
 data is from the 3' end  
 Possible reversed clone: polyT not found.  
 Location/Qualifiers  
 1. .320  
 /organism="Mus musculus"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone\_lib="Stratagene mouse skin (#937313)"  
 /sex="females"  
 /tissue\_type="whole skin"  
 /dev\_stage="11 weeks old"

FEATURES  
 source

Location/Qualifiers  
 1. .320  
 /organism="Mus musculus"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone\_lib="Stratagene mouse skin (#937313)"  
 /sex="females"  
 /tissue\_type="whole skin"  
 /dev\_stage="11 weeks old"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: skin; Vector: pBluescript SK-; Site\_1: EcoRI  
 ; Site\_2: XhoI; Cloned unidirectionally, primer: Oligo  
 dr. Whole skin from 11 week old C57BL/6 female mice.  
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
 adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor  
 sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3"

BASE COUNT 67 a 104 c 85 g 64 t



## ORIGIN

Query Match 1.8%; Score 19; DB 20; Length 320;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ggtggccaggatgctgtg 106  
 |||||  
 Db 248 GGTGGCAGGATGCTGTG 266

## RESULT 24

BF403284  
 LOCUS 358 bp mRNA 28-NOV-2000  
 DEFINITION UI-R-CAL-bly-m-13-0-UI.s1 UI-R-CAL Rattus norvegicus cDNA clone  
 UI-R-CAL-bly-m-13-0-UI 3', mRNA sequence.

ACCESSION BF403284  
 VERSION BF403284  
 KEYWORDS EST.  
 SOURCE BF403284.1 GI:11391259

ORGANISM Norway rat.  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 358)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to identify it as a clone from the  
 normalized corpus-striatum library cDNA Library Preparation: M.B.  
 Soares lab Clone distribution: clones will be available through  
 Research Genetics (www.resgen.com)  
 Seq primer: M13 Forward  
 POLYA=Yes.

## FEATURES

source

Location/Qualifiers

1..358  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-CAL-bly-m-13-0-UI"  
 /clone\_lib="UI-R-CAL"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-CAL  
 library is a subtracted library derived from the following  
 tissues: thalamus, cerebellum, hypothalamus, medulla, pons  
 , midbrain, cerebral cortex, corpus striatum, testis, and  
 hippocampus. For a detailed description of the library  
 from which this clone was derived, please visit our web  
 site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been  
 previously described in (Bonaldo, Lennon and Soares,  
 Genome Research 6:791-806, 1996)  
 TAG\_LIB=UI-R-CAL  
 TAG\_TISSUE=corpus-striatum  
 TAG\_SEQ=CTAGG"

80 a 97 c 78 g 103 t

## BASE COUNT

ORIGIN

Query Match 1.8%; Score 19; DB 148; Length 358;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 tctgcatctcgtcctattt 400  
 |||||  
 Db 132 TCTGCCATCTCGCTATT 150

## RESULT 25

AQ444260/c  
 LOCUS 359 bp DNA GSS 09-JAN-2001  
 DEFINITION GSSTc0303 Trypanosoma cruzi random genomic library Trypanosoma  
 cruzi genomic clone G11D14, DNA sequence.

ACCESSION AQ444260  
 VERSION AQ444260.3 GI:10130822  
 KEYWORDS GSS.

SOURCE Trypanosoma cruzi.  
 ORGANISM Trypanosoma cruzi  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma; Schizotrypanum.

REFERENCE 1 (bases 1 to 359)  
 AUTHORS Agüero,F., Verdun,R., Frasch,A.C.C. and Sanchez,D.O.  
 TITLE A random sequencing approach for the analysis of the trypanosoma  
 cruzi genome: general structure, large gene and repetitive DNA  
 families, and gene discovery

JOURNAL Genome Res. 10 (12), 1996-2005 (2000)  
 MEDLINE 20584849

COMMENT On Sep 14, 2000 this sequence version replaced gi:9378614.  
 Contact: Sanchez D.O.  
 Instituto de Investigaciones Biocnológicas (Univ. Nac. de Gral  
 San Martin)  
 Av. Gral Paz entre Albarcellos y Constituyentes, INTI edificio 24  
 CP(1650) San Martin, Prov. de BS AS. Argentina  
 Tel: 54-11-4580-7255 ext 309  
 Fax: 54-11-4752-9639  
 Email: dsanchez@ib.unsam.edu.ar  
 Sequences were basecalled with phred and vector was masked with  
 crossmatch (see <http://genome.washington.edu>). Sequences were then  
 trimmed from both ends to remove low quality bases and masked  
 vector.

Seq primer: T7

Class: shotgun.

FEATURES Location/Qualifiers

Source 1..359  
 /organism="Trypanosoma cruzi"  
 /strain="CL-Brener"  
 /db\_xref="taxon:5693"  
 /clone="G11D14"  
 /clone\_lib="Trypanosoma cruzi random genomic library"  
 /cell\_type="epimastigote"  
 /note="Vector: pBS(-) (Stratagene); T. cruzi DNA was  
 randomly sheared using a nebulizer and the 1 to 2 Kb range  
 was gel purified and cloned into the dephosphorylated  
 HindIII site of the vector"

84 a 102 c 97 g 76 t

## BASE COUNT

ORIGIN

Query Match 1.8%; Score 19; DB 228; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 cgctgctctctctgctggg 44  
 |||||  
 Db 152 CGCTGCTCTCTCTGCTGG 134

## RESULT 26

BG314065 367 bp mRNA EST 16-APR-2001  
 LOCUS WHE2467\_F07\_K132S Triticum monococcum early reproductive apex cDNA  
 DEFINITION library Triticum monococcum cDNA clone WHE2467\_F07\_K13, mRNA

```

sequence.
ACCESSION   BG314065
VERSION     BG314065.1  GI:13115868
KEYWORDS    EST.
SOURCE      Triticum monococcum.
ORGANISM    Triticum monococcum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1 (bases 1 to 367)
AUTHORS    Anderson,O.D., Chao,S., Dubcovsky,J., Echenique,V., Han,P.S., Hsia
            ,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L.,
            Stanova,B. and Tong,J.C.
TITLE       The structure and function of the expressed portion of the wheat
            genomes - Early reproductive apex cDNA library from Triticum
            monococcum
JOURNAL     Unpublished (2001)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agricultural Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oanderson@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES    Location/Qualifiers
             1..367
                /organism="Triticum monococcum"
                /cultivar="DV92"
                /db_xref="taxon:4568"
                /clone="WHE2467_F07_K13"
                /clone_lib="Triticum monococcum early reproductive apex
                cDNA library"
                /tissue_type="Early reproductive apex"
                /dev_stage="Seven week-old plants"
                /lab_host="E. coli XL0LR"
                /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                Site_1: EcoRI; Site_2: XhoI; The tissue, total RNA, and
                poly(A) RNA were prepared from apex at double-ridge stage
                to terminal-spikelet stage during transition from
                vegetative state to flower state, a cDNA library was made,
                and the cDNA clones were in vivo excised at the
                University of California, Davis (V. Echenique, B. Stanova
                , J. Dubcovsky). Plasmid DNA preparations and DNA
                sequencing were performed in the OD Anderson lab (all
                other authors)."
BASE COUNT  59 a 109 c 107 g 92 t
ORIGIN

```

```

Query Match      1.8%; Score 19; DB 152; Length 367;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 873 ctctgtgctcttggaacct 891
|||||
Db 207 CTCCTGTGCCTTTGGACCT 225

RESULT 27
LOCUS      BF523355          372 bp      mRNA      EST      11-DEC-2000
DEFINITION UI-R-G0-ug-d-03-0-UI.r1 UI-R-G0 Rattus norvegicus cDNA clone
            UI-R-G0-ug-d-03-0-UI 5', mRNA sequence.
ACCESSION  BF523355
VERSION    BF523355.1  GI:11631322
KEYWORDS   EST.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

```

```

Rattus.
1 (bases 1 to 372)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
CDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LLNL (info@image.llnl.gov). IMAGE ID= 1794012 The following
repetitive elements were found in this cDNA sequence: 144-173,
>AT-rich#Low_complexity
Seq primer: M13 Forward.
            Location/Qualifiers
             1..372
                /organism="Rattus norvegicus"
                /strain="Sprague-Dawley"
                /db_xref="taxon:10116"
                /clone="UI-R-G0-ug-d-03-0-UI"
                /clone_lib="UI-R-G0"
                /dev_stage="adult"
                /lab_host="DH10B (Life Technologies)"
                /note="Vector: pT7m3D-Pac (Pharmacia) with a modified
                polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-G0
                library is a normalized library constructed from a
                mixture of rat tissues (nodose ganglia, dorsal root
                ganglia, and trigeminal ganglia). The tag is a string of
                6 nucleotides present between the Not I site and the
                oligo-dT track. The library was constructed as described
                by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
                , 1996."
BASE COUNT  112 a 91 c 77 g 92 t
ORIGIN

```

```

Query Match      1.8%; Score 19; DB 150; Length 372;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 404 tgccagtggtcacaagca 422
|||||
Db 282 TGCCAGTGTCAAAAGCA 300

RESULT 28
LOCUS      AQ0909026/c      401 bp      DNA      GSS      09-JAN-2001
DEFINITION GSStc09872 Trypanosoma cruzi random genomic library Trypanosoma
            cruzi genomic clone G4D14, DNA sequence.
ACCESSION  AQ0909026
VERSION    AQ0909026.1  GI:6489356
KEYWORDS   GSS.
SOURCE     Trypanosoma cruzi.
            Trypanosoma cruzi
            Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
            1 (bases 1 to 401)
            Aquero,F., Verdun,R., Frasch,A.C.C. and Sanchez,D.O.
            A random sequencing approach for the analysis of the trypanosoma
            cruzi genome: general structure, large gene and repetitive DNA
            families, and gene discovery
            Genome Res. 10 (12), 1996-2005 (2000)
            20568489
            Contact: Sanchez D.O.
            Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral

```

San Martin)  
 Av. Gral Paz entre Albarcellos y Constituyentes, INTI edificio 24  
 CP(1650) San Martin, Prov. de BS AS. Argentina  
 Tel.: 54-11-4580-7255 ext 309  
 Fax: 54-11-4752-9639  
 Email: dsanchez@lib.unsam.edu.ar  
 Seq primer: T7  
 Class: shotgun.

# FEATURES

source  
 1. 401  
 /organism="Trypanosoma cruzi"  
 /strain="CL-Brener"  
 /db\_xref="taxon:5693"  
 /clone="G4D14"  
 /clone\_lib="Trypanosoma cruzi random genomic library"  
 /cell\_type="epimastigote"  
 /note="Vector: pBS(-) (Stratagene); T. cruzi DNA was  
 randomly sheared using a nebulizer and the 1 to 2 Kb range  
 was gel purified and cloned into the dephosphorylated  
 HincII site of the vector"

BASE COUNT 83 a 117 c 106 g 92 t 3 others  
 ORIGIN  
 Query Match 1.8%; Score 19; DB 235; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 cgcgtctctcttctgtggg 44  
 |||||  
 Db 143 CGCTGCTCTCTTCTGCTGGG 125

# RESULT 29

AQ178480 417 bp DNA GSS 17-OCT-1998  
 LOCUS HS-2222\_B1\_E01\_MF CIT Approved Human Genomic Sperm Library D Homo  
 DEFINITION sapiens genomic clone plate=2222 Col-1 Row-J, DNA sequence.  
 ACCESSION AQ178480  
 VERSION AQ178480.1 GI:3575847  
 KEYWORDS GSS.  
 SOURCE human.

# ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 417)  
 MAHAIKAS, G.G., WALLACE, J.C., SMITH, K., SWARTZELL, S., HOLZMAN, T.,  
 KELLER, A., SHAKER, R., FURLONG, J., YOUNG, J., ZHAO, S., ADAMS, M.D. and  
 HOOD, L.

Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

# JOURNAL

MEDLINE  
 COMMENT  
 Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallaceu.washington.edu  
 Sequence Tagged Connector  
 Plate: 2222 row: J column: 1  
 Class: BAC ends  
 High quality sequence stop: 417.  
 Location/Qualifiers  
 1. 417  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate-2222 Col-1 Row-J"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in  
 E-Coli DH108"

# FEATURES

source  
 1. 417  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate-2222 Col-1 Row-J"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in  
 E-Coli DH108"

BASE COUNT 147 a 85 c 76 g 104 t 5 others  
 ORIGIN

Query Match 1.8%; Score 19; DB 225; Length 417;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 671 tgcctgtcacattgatgg 689  
 |||||  
 Db 139 TGTCTGTCTCATTTGATGG 157

# RESULT 30

A1562186 418 bp mRNA EST 25-MAR-1999  
 LOCUS VW74e07.xl Stratagene mouse heart (#937316) Mus musculus CDNA clone  
 DEFINITION IMAGE:1260708 3' similar to SW:RGSC\_RAT 008774 REGULATOR OF  
 G-PROTEIN SIGNALLING 12 ;, mRNA sequence.

ACCESSION A1562186  
 VERSION A1562186.1 GI:4513531  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 418)  
 MARA, M., HILLIER, L., KUCABA, T., MARTIN, J., BECK, C., WYLIE, T.,  
 UNDERWOOD, K., STEPTOE, M., THEISING, B., ALLEN, M., BOWERS, Y., PERSON  
 B., SWALLIER, T., GIBBONS, M., PAPE, D., HARVEY, N., SCHURK, R., RITTER  
 E., KOHN, S., SHIN, T., JACKSON, Y., CARDENAS, M., MCCANN, R.,  
 WATERSTON, R. and WILLSON, R.

The WashU-NCI Mouse EST Project 1999  
 Unpublished (1999)

CONTACT: Marra M/WashU-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:663260

This clone was previously sequenced on the 5' end only, this new  
 data is from the 3' end  
 Possible reversed clone: similarity on wrong strand  
 High quality sequence stop: 233.  
 Location/Qualifiers  
 1. 418  
 /organism="Mus musculus"  
 /strain="NIH/Swiss"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1260708"  
 /clone\_lib="Stratagene mouse heart (#937316)"  
 /sex="pooled"  
 /tissue\_type="heart"  
 /dev\_stage="13 day embryos"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: heart; Vector: pBluescript SK-; Site: 1;  
 EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer:  
 Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.  
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
 adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor  
 sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' "

# FEATURES

source  
 1. 418  
 /organism="Mus musculus"  
 /strain="NIH/Swiss"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1260708"  
 /clone\_lib="Stratagene mouse heart (#937316)"  
 /sex="pooled"  
 /tissue\_type="heart"  
 /dev\_stage="13 day embryos"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: heart; Vector: pBluescript SK-; Site: 1;  
 EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer:  
 Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.  
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
 adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor  
 sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' "

# BASE COUNT

93 a 129 c 114 g 82 t  
 ORIGIN  
 Query Match 1.8%; Score 19; DB 22; Length 418;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ggtgcccagatgctgctg 106

```

|||||
Db 230 GGTGGCCAGGATGCTGCTG 248

RESULT 31
AA855540/c
LOCUS
DEFINITION
vW74e07_r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
IMAGE:1260708 5', similar to SW:RGSC_RAT 008774 REGULATOR OF
G-PROTEIN SIGNALLING 12 ;, mRNA sequence.

ACCESSION
AA855540
VERSION
AA855540.1 GI:2943078
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 438)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:863260
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 428.
Location/Qualifiers
1. .438
/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone="IMAGE:1260708"
/clone_lib="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: heart; Vector: pBluescript SK-; Site:1:
ECORI; Site:2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 Kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'
1 others
BASE COUNT 95 a 114 c 132 g 96 t
ORIGIN

Query Match 1.8%; Score 19; DB 12; Length 438;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ggtgcccagatgctgctg 106
|||||
Db 398 GGTGGCCAGGATGCTGCTG 380

RESULT 32
TA62C120/c
LOCUS
DEFINITION
T. brucei sheared genomic DNA clone 62c12, reverse sequence,
genomic survey sequence.
ACCESSION
AL464602

```

```

VERSION
KEYWORDS
SOURCE
ORGANISM
AL464602.1 GI:11859303
GSS.
Trypanosoma brucei.
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 453)
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nhl@sanger.ac.uk
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.
Location/Qualifiers
1. .463
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="62c12"
BASE COUNT 101 a 122 c 94 g 146 t
ORIGIN

Query Match 1.8%; Score 19; DB 258; Length 463;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 604 gacagatttgctgctg 622
|||||
Db 75 GACAAGATTGTGCTGCTG 57

RESULT 33
BE133527
LOCUS
DEFINITION
ug34b11.y1 Soares_NMPu Mus musculus cDNA clone IMAGE:1533501 5'
similar to TR:088383 088383 RGS12 PDZ-LESS VARIANT. ;, mRNA
sequence.
ACCESSION
BE133527
VERSION
BE133527.1 GI:8596027
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 471)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:949601
Possible reverse clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 411.
Location/Qualifiers

```

```

source
1. .471
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1533501"
/clone_lib="Soares_NMPu"
/sex="female"
/dev_stage="adult"
/lab_host="DH108"
/note="Organ: uterus; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; 1st strand cDNA was prepared from
pregnant mouse uterus, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      90 a 139 c 122 g 120 t
ORIGIN

Query Match      1.8%; Score 19; DB 163; Length 471;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ggtggccaggatgctgctg 106
      |||||||
Db 384 GGTGGCAGGATGCTGCTG 402

RESULT 34
AQ413533/c
LOCUS
DEFINITION      AQ413533      506 bp      DNA      GSS      23-MAR-1999
RPCI-11-169C8.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-169C8,
DNA sequence.
ACCESSION      AQ413533
VERSION
KEYWORDS
SOURCE
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. (bases 1 to 506)
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
Unpublished (1997)
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbsetigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers
source
1. .506
/organism="Homo sapiens"
/db_xref="GDB:7564567"
/db_xref="taxon:9606"
/clone="RPCI-11-169C8"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"

```

```

BASE COUNT      148 a 98 c 88 g 171 t 1 others
ORIGIN

Query Match      1.8%; Score 19; DB 228; Length 506;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 609 gatttgctggtgataact 627
      |||||||
Db 276 GATTGTGCTGGTGACTACT 258

RESULT 35
AQ646668/c
LOCUS
DEFINITION      AQ646668      515 bp      DNA      GSS      08-JUL-1999
RPCI93-DpnII-25E14.TJ RPCI93- Trypanosoma brucei genomic clone
RPCI93-DpnII-25E14, DNA sequence.
ACCESSION      AQ646668
VERSION
KEYWORDS
SOURCE
ORGANISM      Trypanosoma brucei.
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE
1. (bases 1 to 515)
El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
Gerrard,C., Leech,V., de Jong,P., Ullu,E., Meiville,S., Donelson,J.
, Fraser,C. and Adams,M.
Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPCI-93
Library for gene discovery and sequence-ready map construction
Unpublished (1999)
Other_GSSs: RPCI93-DpnII-25E14.TV
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones and high density filters may be purchased from BACPAC
Resources (http://bacpac.med.buffalo.edu. BAC end sequences search
page: http://www.tigr.org/tdb/mbd/tbdb/.
Seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers
source
1. .515
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="RPCI93-DpnII-25E14"
/clone_lib="RPCI93-DpnII"
/note="Vector: pBACe3.6; Site_1: Bam HI; Site_2: Bam HI;
Constructed for The Institute for Genomic Research by
Bohui Zhao in Pieter de Jong's laboratory (Roswell Park
Cancer Institute, Buffalo, NY). Briefly, Trypanosoma
brucei TREU927/4 GUTat 10.1 agarose embedded DNA was
partially digested with a combination of Eco RI and Eco RI
methylase (RPCI93-EcoRI segment) or Dpn II (RPCI93-DpnII
segment). High molecular weight fragments were ligated in
pBACe3.6 vector digested with Eco RI or Bam HI,
respectively. The average insert size is 141 Kb. Total
coverage (both segments): > 90 x the haploid
non-minichromosomal genome."
BASE COUNT      110 a 133 c 149 g 123 t
ORIGIN

Query Match      1.8%; Score 19; DB 231; Length 515;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 434 caccctttgttggtgac 452

```

```
Db 117 CACCCTTTGTTGGGTGAC 99
|||||
RESULT 36
BF756411 523 bp mRNA EST 12-JAN-2001
LOCUS PM4-CT0113-301000-001-d06 CT0113 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF756411
ACCESSION BF756411
VERSION BF756411.1 GI:12104311
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 523)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&t2=PM4-CT0113-
301000-001-d06&t3=2000-10-30&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 23
High quality sequence stop: 523.
FEATURES
source
1..523
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0113"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 159 a 98 c 134 g 131 t 1 others
ORIGIN
Query Match 1.8%; Score 19; DB 169; Length 523;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 456 atggggaaagttaaggaa 474
|||||
Db 138 ATGGGGAAGAGTTAANGGAA 156
|||||
RESULT 37
AW650509 527 bp mRNA EST 04-APR-2000
LOCUS EST328963 tomato germinating seedlings, TAMU Lycopersicon
DEFINITION EST328963
ACCESSION AW650509
SOURCE esculentum cDNA clone cLE113A10 5', mRNA sequence.
AW650509
```

```
VERSION AW650509.1 GI:7411747
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; eusterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 527)
AUTHORS Alcalá,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley
,S.D.
TITLE Generation of ESTs from germinating tomato seed
JOURNAL Unpublished (2000)
COMMENT Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
source
1..527
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone_lib="cLE113A10"
/clone_lib="tomato germinating seedlings, TAMU"
/tissue_type="whole seedlings"
/dev_stage="7 days post imbibition"
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; 7 days post imbibition on water-agar. Mixed stage
whole germinating seedlings from seed coat emergence up
to two centimeters in seeds not showing obvious signs of
germination were discarded."
BASE COUNT 172 a 98 c 88 g 169 t
ORIGIN
Query Match 1.8%; Score 19; DB 119; Length 527;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 781 tggattaatgccactattt 799
|||||
Db 109 TGGATTAAATGCCACTATT 127
|||||
RESULT 38
BE030483/c 552 bp mRNA EST 09-JUL-2000
LOCUS 128575 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
DEFINITION BE030483
ACCESSION BE030483
VERSION BE030483.1 GI:8325492
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 552)
AUTHORS Fahrenkrug,S.C., Frenking,B.A., Rohrer,G.A., Smith,T.P.L., Casae,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for
JOURNAL ESR discovery in swine
COMMENT Unpublished (2000)
Contact: Smith Tpl
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
```

Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.

PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCAGTCAGCAGC  
Plate: 62 row: M column: 11  
Seq primer: ATTAGTGTGACACTATAG.

## FEATURES

source  
1. .552  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone\_lib="MARC 1P1G"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from day 11, 13, 15, 20,  
and 30 embryos."

BASE COUNT 192 a 127 c 150 g 83 t  
ORIGIN

Query Match 1.8%; Score 19; DB 162; Length 552;

Best Local Similarity 100.0%; Pred. No. 41;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 tgcgtgtgcaggcgctgg 117  
|||||  
DB 280 TGCTGCTGCAGGCGCTGG 262

## RESULT 39

LOCUS AI644407 563 bp mRNA EST 29-APR-1999  
DEFINITION vi92a05.x1 Stratagene mouse heart (#937316) Mus musculus cDNA clone  
IMAGE:919664 3' similar to TR:088383 O88383 RGS12 PDZ-LESS VARIANT.  
; mRNA sequence.

ACCESSION AI644407  
VERSION AI644407.1 GI:4722882  
KEYWORDS EST.  
SOURCE house mouse.

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 563)  
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person  
B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter  
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
Waterston, R., and Wilson, R.

The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)

JOURNAL Contact: Marra M/WashU-NCI Mouse EST Project 1999  
COMMENT Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:531880

This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end  
Possible reversed clone: similarity on wrong strand  
Possible reversed clone: polyT not found  
High quality sequence stop: 438.

## FEATURES

source  
1. .563  
/organism="Mus musculus"  
/strain="NIH/Swiss"  
/db\_xref="taxon:10090"  
/clone="IMAGE:919664"

/clone\_lib="Stratagene mouse heart (#937316)"  
/sex="pooled"  
/tissue\_type="heart"  
/dev\_stage="13 day embryos"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: heart; Vector: pBluescript SK-; Site\_1:  
ECORI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.  
Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'  
adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor  
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3"

BASE COUNT 121 a 175 c 150 g 116 t 1 others  
ORIGIN

Query Match 1.8%; Score 19; DB 23; Length 563;

Best Local Similarity 100.0%; Pred. No. 41;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ggtggccaggatgctgctg 106  
|||||  
DB 248 GGTGGCCAGGATGCTGCTG 266

## RESULT 40

LOCUS AQ307958 590 bp DNA GSS 22-DEC-1998  
DEFINITION CIT-HSP-2385023.TRB CIT-HSP Homo sapiens genomic clone 2385023, DNA  
sequence.

ACCESSION AQ307958  
VERSION AQ307958.1 GI:4040310  
KEYWORDS GSS.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 590)  
Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,  
Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and  
Venter, J.C.

Use of a random human BAC End Sequence Database for Sequence-Ready  
Map Building  
Unpublished (1998)

JOURNAL Other GSS: CIT-HSP-2385023.TR CIT-HSP-2385023.TF  
COMMENT Contact: Shaving Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics

The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.

Seq primer: M13 Reverse  
Class: BAC ends.

## FEATURES

source  
Location/Qualifiers  
1. .590  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2385023"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2:  
HindIII"

BASE COUNT 191 a 108 c 107 g 184 t  
ORIGIN

Query Match 1.8%; Score 19; DB 226; Length 590;

Best Local Similarity 100.0%; Pred. No. 41;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 609 gatttgctgtgataact 627  
 |||||  
 Db 164 GATTGTGCTGTGATACT 182

RESULT 41  
 AQ197598  
 LOCUS  
 DEFINITION CIT-HSP-2384023.TR CIT-HSP Homo sapiens genomic clone 2384023, DNA sequence.  
 ACCESSION AQ197598  
 VERSION AQ197598.1 GI:3604960  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 596)  
 AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.  
 TITLE Use of a random human BAC End Sequence Database for Sequence-Ready Map Building  
 JOURNAL Unpublished (1998)  
 COMMENT Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC end search page:  
 http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
 Seq primer: M13 Reverse  
 Class: BAC ends.

FEATURES  
 Location/Qualifiers  
 1..596  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="2384023"  
 /clone\_lib="CIT-HSP"  
 /sex="Male"  
 /cell\_type="Sperm"  
 /note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2: HindIII"

BASE COUNT 190 a 107 c 110 g 189 t  
 ORIGIN

Query Match 1.8%; Score 19; DB 225; Length 596;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 609 gatttgctgtgataact 627  
 |||||  
 Db 164 GATTGTGCTGTGATACT 182

RESULT 42  
 AQ923548  
 LOCUS  
 DEFINITION AQ923548.1 TV RPCI-23 Mus musculus genomic clone RPCI-23-299K17, DNA sequence.  
 ACCESSION AQ923548  
 VERSION AQ923548.1 GI:6612551  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 621)

AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.  
 TITLE Mouse BAC End Sequences from Library RPCI-23  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page:  
 http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 299 row: K column: 17  
 Seq primer: T7  
 Class: BAC ends.

FEATURES  
 Location/Qualifiers  
 1..621  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-299K17"  
 /clone\_lib="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methyase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 222 a 118 c 156 g 125 t  
 ORIGIN

Query Match 1.8%; Score 19; DB 235; Length 621;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 704 caggagtagtaagctggg 722  
 |||||  
 Db 555 CAGGAGTAGTAGCTGGG 573

RESULT 43  
 AQ727580  
 LOCUS  
 DEFINITION HS\_5456.A2\_H09\_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1032 Col=18 Row=O, DNA sequence.  
 ACCESSION AQ727580  
 VERSION AQ727580.1 GI:5487249  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 627)  
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 MEDLINE 99380589  
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington



401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm)) or from Resear h Genetics (<http://info@resgen.com>). BAC end Web Server: <http://www.htsc.washington.edu>  
 Plate: 1032 row: 0 column: 18  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence stop: 627.  
 Location/Qualifiers  
 1. .627  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="RPCI-11 Human Male BAC Library"  
 /sex="male"  
 /note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"  
 BASE COUNT 201 a 148 c 118 g 144 t 16 others  
 ORIGIN

Query Match 1.8%; Score 19; DB 232; Length 627;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 773 accaaaaatggattaatgc 791  
 ||||||||||||||||  
 Db 403 ACCAAAATGGATTATGC 421

RESULT 44  
 AZ913822/c  
 LOCUS  
 DEFINITION AZ913822 664 bp DNA GSS 05-MAR-2001  
 RPCI-24-176612-TV RPCI-24 Mus musculus genomic clone RPCI-24-176612  
 , DNA sequence.  
 ACCESSION AZ913822  
 VERSION AZ913822.1 GI:13232767  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 664)  
 Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.  
 Mouse BAC End Sequences from Library RPCI-24  
 Unpublished (1999)  
 CONTACT: Shaving Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdjong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 176 row: 6 column: 12  
 Seq primer: T7  
 Class: BAC ends.  
 Location/Qualifiers

FEATURES  
 source

1. .664  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-24-176612"  
 /clone\_lib="RPCI-24"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /note="Vector: pTARBAC1; Site\_1: BamHI; Site\_2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."  
 BASE COUNT 277 a 130 c 96 g 161 t  
 ORIGIN

Query Match 1.8%; Score 19; DB 251; Length 664;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 607 aagatttgctgtgata 625  
 ||||||||||||||||  
 Db 132 AAGATTGTGCTGTGATA 114

RESULT 45  
 BF167775/c  
 LOCUS  
 DEFINITION BF167775 697 bp mRNA EST 30-OCT-2000  
 601773578F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3994146 5', mRNA sequence.  
 ACCESSION BF167775  
 VERSION BF167775.1 GI:11048127  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 697)  
 NIH-MGC <http://mgc.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: cgabbs@mail.nih.gov  
 Tissue procurement: Gilbert Smith, Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LLAM9211 row: 1 column: 19  
 High quality sequence stop: 684.  
 Location/Qualifiers  
 1. .697  
 /organism="Mus musculus"  
 /strain="C57BL/6J (fetal)"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3994146"  
 /clone\_lib="NCI\_CGAP\_Lu29"  
 /tissue\_type="spontaneous tumor, metastatic to mammary. Stem cell origin."  
 /lab\_host="DH10B"  
 /note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"  
 BASE COUNT 161 a 163 c 222 g 149 t  
 ORIGIN

Query Match 1.8%; Score 19; DB 145; Length 697;  
 Best Local Similarity 100.0%; Pred. No. 42;

---

Matches	19;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	88	ggtggccaggatgctg	106						
Db	463	GGTGGCCAGGATGCTG	445						

Search completed: November 22, 2001, 01:25:25  
Job time: 3939 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2001, 23:31:50 ; Search time 1439.34 Seconds  
(without alignments)  
11187.021 Million cell updates/sec

Title: US-09-854-844-1  
Perfect score: 1041  
Sequence: 1 atgggcctgctgctgtgc.....ttattacaatttgaatga 1041

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl:\*
- 1: gb\_ba1:\*
  - 2: gb\_ba2:\*
  - 3: gb\_ba3:\*
  - 4: gb\_in1:\*
  - 5: gb\_in2:\*
  - 6: gb\_in3:\*
  - 7: gb\_om:\*
  - 8: gb\_ov:\*
  - 9: gb\_pat1:\*
  - 10: gb\_pat2:\*
  - 11: gb\_ph:\*
  - 12: gb\_pl1:\*
  - 13: gb\_pl2:\*
  - 14: gb\_pl3:\*
  - 15: gb\_pl4:\*
  - 16: em\_ba1:\*
  - 17: em\_ba2:\*
  - 18: em\_fun:\*
  - 19: em\_htgo\_hum:\*
  - 20: em\_htgo\_inv:\*
  - 21: em\_htgo\_rod:\*
  - 22: em\_htg\_hum1:\*
  - 23: em\_htg\_hum2:\*
  - 24: em\_htg\_hum3:\*
  - 25: em\_htg\_hum4:\*
  - 26: em\_htg\_hum5:\*
  - 27: em\_htg\_hum6:\*
  - 28: em\_htg\_hum7:\*
  - 29: em\_htg\_hum8:\*
  - 30: em\_htg\_inv1:\*
  - 31: em\_htg\_inv2:\*
  - 32: em\_htg\_other:\*
  - 33: em\_htg\_rod:\*
  - 34: em\_hum1:\*
  - 35: em\_hum2:\*
  - 36: em\_hum3:\*
  - 37: em\_hum4:\*
  - 38: em\_hum5:\*
  - 39: em\_hum6:\*
  - 40: em\_hum7:\*
  - 41: em\_in:\*
  - 42: em\_om:\*
  - 43: em\_or:\*

- 44: em\_ov:\*
- 45: em\_pat:\*
- 46: em\_ph:\*
- 47: em\_pl:\*
- 48: em\_ro:\*
- 49: em\_sts:\*
- 50: em\_sy:\*
- 51: em\_un:\*
- 52: em\_vl:\*
- 53: gb\_sts1:\*
- 54: gb\_sts2:\*
- 55: gb\_sts3:\*
- 56: gb\_sy:\*
- 57: gb\_un:\*
- 58: gb\_vl:\*
- 59: gb\_vl2:\*
- 60: gb\_htg1:\*
- 61: gb\_htg2:\*
- 62: gb\_htg3:\*
- 63: gb\_htg4:\*
- 64: gb\_htg5:\*
- 65: gb\_htg6:\*
- 66: gb\_htg7:\*
- 67: gb\_htg8:\*
- 68: gb\_htg9:\*
- 69: gb\_htg10:\*
- 70: gb\_htg11:\*
- 71: gb\_htg12:\*
- 72: gb\_htg13:\*
- 73: gb\_htg14:\*
- 74: gb\_htg15:\*
- 75: gb\_htg16:\*
- 76: gb\_htg17:\*
- 77: gb\_htg18:\*
- 78: gb\_htg19:\*
- 79: gb\_htg20:\*
- 80: gb\_htg21:\*
- 81: gb\_htg22:\*
- 82: gb\_htg23:\*
- 83: gb\_htg24:\*
- 84: gb\_htg25:\*
- 85: gb\_pr1:\*
- 86: gb\_pr2:\*
- 87: gb\_pr3:\*
- 88: gb\_pr4:\*
- 89: gb\_pr5:\*
- 90: gb\_pr6:\*
- 91: gb\_pr7:\*
- 92: gb\_pr8:\*
- 93: gb\_pr9:\*
- 94: gb\_rol:\*
- 95: gb\_rod:\*
- 96: gb\_in4:\*
- 97: gb\_pr10:\*
- 98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	391.4	37.6	178181	77	AC084420	AC084420 Homo sapi
2	391.4	37.6	180155	70	AC026556	AC026556 Homo sapi
3	266.2	25.6	178181	77	AC084420	AC084420 Homo sapi
4	151	14.5	2078	8	AB018694	AB018694 Xenopus l
5	132.8	12.8	942	10	AX082979	AX082979 Sequence
6	129.6	12.4	1850	8	AB038406	AB038406 Xenopus l
7	129.4	12.4	1055	94	AY005145	AY005145 Mus muscu
8	129.4	12.4	1095	94	AF176209	AF176209 Mus muscu

```

9 118.6 11.4 942 10 E21865
10 118.6 11.4 1082 10 E21866
11 118.6 11.4 1082 10 E21867
12 118.6 11.4 1082 85 AB031329 Homo sapi
13 117 11.2 1071 85 AB031330
14 117 11.2 1077 88 AF058300 Homo sapi
15 115.8 11.1 1013 93 HGA306593
16 107.2 10.3 945 10 E21852
17 107.2 10.3 1085 10 E21853
18 107.2 10.3 1085 10 E21854
19 103.2 9.9 1797 94 AF188613
20 100 9.6 1743 94 BC003851 Mus muscu
21 98.6 9.5 936 10 AX082992 Sequence
22 98.2 9.4 1431 94 MMU243866
23 97.4 9.4 1796 10 AX098193
24 97.4 9.4 1809 91 BC001462
25 97.4 9.4 1834 97 HUMPROS
26 97.4 9.4 1835 10 AX098215
27 96.2 9.2 1154 10 AX082977
28 91.2 8.8 1090 94 MUSMMP
29 91.2 8.8 1097 9 AR080456
30 91.2 8.8 1097 9 RATRMCT
31 91.2 8.8 1108 9 AR080464
32 91.2 8.8 1108 94 MUSPROT6B
33 89.6 8.6 1103 95 RNU67909
34 88 8.5 1321 94 AB010778
35 86.6 8.3 1122 94 AF175523
36 84.6 8.1 2208 94 AB017638
37 83 8.0 1864 94 AF202076
38 82.6 7.9 1259 95 RRA5642
39 81.8 7.9 736 94 MMU42405
40 81.8 7.9 1031 9 AR080454
41 81.8 7.9 1031 94 MUSPROTEA
42 80.8 7.8 1235 95 RNU67910
43 77.6 7.5 1339 8 AF029404
44 76.4 7.3 1221 9 AX014344
45 75.6 7.3 995 7 DOGMCTRA

```

## ALIGNMENTS

```

RESULT 1
AC084420/c
LOCUS AC084420 178181 bp DNA HTG 11-DEC-2000
DEFINITION Homo sapiens chromosome RP11-415A13, WORKING DRAFT
SEQUENCE, 32 unordered pieces.
AC084420
VERSION AC084420.3 GI:11612633
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 178181)
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 178181)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Dec 11, 2000 this sequence version replaced gi:11136874.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0415A13
----- Summary Statistics -----

```

```

Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer Et; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 160877 bases at least Q40
Consensus quality: 166793 bases at least Q30
Consensus quality: 169458 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 175081; sum-of-contigs
Quality coverage: 3.40 in Q20 bases; agarose-fp
Quality coverage: 3.71 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1401: contig of 1401 bp in length
* 1402 1501: gap of unknown length
* 1502 2671: contig of 1170 bp in length
* 2672 2771: gap of unknown length
* 2772 4499: contig of 1728 bp in length
* 4500 4599: gap of unknown length
* 4600 6309: contig of 1710 bp in length
* 6310 6409: gap of unknown length
* 6410 8601: contig of 2192 bp in length
* 8602 8701: gap of unknown length
* 8702 10415: contig of 1714 bp in length
* 10416 10515: gap of unknown length
* 10516 13064: contig of 2549 bp in length
* 13065 13164: gap of unknown length
* 13165 16156: contig of 2992 bp in length
* 16157 16256: gap of unknown length
* 16257 18638: contig of 2382 bp in length
* 18639 18738: gap of unknown length
* 18739 21462: contig of 2724 bp in length
* 21463 21562: gap of unknown length
* 21563 24282: contig of 2720 bp in length
* 24283 24382: gap of unknown length
* 24383 28368: contig of 3986 bp in length
* 28369 28468: gap of unknown length
* 28469 32249: contig of 3781 bp in length
* 32250 32349: gap of unknown length
* 32350 36555: contig of 4206 bp in length
* 36556 36655: gap of unknown length
* 36656 42852: contig of 6197 bp in length
* 42853 42952: gap of unknown length
* 42953 47243: contig of 4291 bp in length
* 47244 47343: gap of unknown length
* 47344 53015: contig of 5672 bp in length
* 53016 53116: gap of unknown length
* 53117 59823: contig of 6707 bp in length
* 59824 59923: gap of unknown length
* 59924 65441: contig of 5419 bp in length
* 65442 65441: gap of unknown length
* 65442 70898: contig of 5457 bp in length
* 70899 70998: gap of unknown length
* 70999 76921: contig of 5923 bp in length
* 76922 77021: gap of unknown length
* 77022 85604: contig of 8583 bp in length
* 85605 85704: gap of unknown length
* 85705 90872: contig of 5168 bp in length
* 90873 90972: gap of unknown length
* 90973 97315: contig of 6343 bp in length
* 97316 97415: gap of unknown length
* 97416 105928: contig of 8513 bp in length
* 105929 106028: gap of unknown length
* 106029 116253: contig of 10125 bp in length
* 116254 116253: gap of unknown length

```





misc\_feature 26627..31028 /note="assembly\_fragment"  
misc\_feature 31129..38352 /note="assembly\_fragment"  
misc\_feature 38453..43015 /note="assembly\_fragment"  
misc\_feature 43116..49226 /note="assembly\_fragment"  
misc\_feature 49327..54582 /note="assembly\_fragment"  
misc\_feature 54683..60160 /note="assembly\_fragment"  
misc\_feature 60261..66143 /note="assembly\_fragment"  
misc\_feature 66244..74231 /note="assembly\_fragment"  
misc\_feature 74332..80835 /note="assembly\_fragment"  
misc\_feature clone\_end:SP6  
vector\_side:right  
80936..91055 /note="assembly\_fragment"  
91156..101763 /note="assembly\_fragment"  
101864..113795 /note="assembly\_fragment"  
113896..125368 /note="assembly\_fragment"  
125469..145206 /note="assembly\_fragment"  
145307..162142 /note="assembly\_fragment"  
Query Match 37.6%; Score 391.4; DB 70; Length 180155;  
Best Local Similarity 95.7%; Pred. No. 5.3e-111;  
Matches 401; Conservative 1; Mismatches 17; Indels 0; Gaps 0;  
QY 623 atactcaaacatgaaggatagttgcaagggtgattcttgaggggcctctgtgtgcaca 682  
DB 116298 ATGCTATTACCTTTTCATTTCTCTTAGGTTGATTTCTGGAGGGCTCTGTCTGTGCACA 116357  
QY 683 tigtatgtgtatggatccacagagagtagtagtgcgtgggattagaattgttaaatctc 742  
DB 116358 TTGATGTTGTATGGATCCACAGAGAGTAGTAGTGGGGATAGAAATGTTGTTAAATCTC 116417  
QY 743 ttcttgaggatctacaccaaataatgtaatactacacaaatggattatgcactatttcaa 802  
DB 116418 TTCTTGAGCTACACCAATGTAATCTACTACCAAAATGGATTATGCGCACTATTTCAA 116477  
QY 803 gagcaacaatctagactctctgtactcttcttccctattgtccctactctctctgtctc 862  
DB 116478 GAGCCAAACAATAGACTTCTGTACTTCTGTTCTTCTTCTTCTTCTTCTTCTGCTC 116537  
QY 863 tctgtgtccctcctgtccttggacttaacacacatacacagagtaggacctgtactgt 922  
DB 116538 TCTTGCGTCCCTCTCTGTGCTTTGGACCTTAACACATATACACAGAGTAGGACTGTAGCTG 116597  
QY 923 aagctgttcttgatcacacagggtcgggaagaagaatgcatggagatttagtccaggggca 982  
DB 116598 AAGCTGTTCTTGATACACAGGCTGGGAGAGAGATGATGAGATTAGTCCAGGGGCA 116657  
QY 983 gagaactcagagagcactgctaaacctgggtgaactttattacaaatttgaaatga 1041  
DB 116658 GATAACTCACAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 116716  
RESULT 3  
AC084420 178181 bp DNA HTG 11-DEC-2000  
LOCUS AC084420 Homo sapiens chromosome RP11-11 clone RP11-415A13, WORKING DRAFT  
DEFINITION AC084420 Homo sapiens chromosome RP11-11 clone RP11-415A13, WORKING DRAFT  
SEQUENCE AC084420 SEQUENCE, 32 unordered pieces.  
ACCESSION AC084420  
VERSION AC084420.3 GI:11612633

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 178181)  
Waterston,R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 178181)  
Waterston,R.H.  
Direct Submission  
Submitted (03-NOV-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Dec 11, 2000 this sequence version replaced gi:11136874.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H\_NH0415A13  
----- Summary Statistics -----  
Sequencing vector: M13; 100%  
Chemistry: Dye-primer ET; 100% of reads  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 160877 bases at least Q40  
Consensus quality: 166793 bases at least Q30  
Consensus quality: 169458 bases at least Q20  
Insert size: 188000; agarose-fp  
Quality coverage: 3.40 in Q20 bases; agarose-fp  
Quality coverage: 3.71 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 32 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1401: contig of 1401 bp in length  
1402 1501: gap of unknown length  
1502 2871: contig of 1170 bp in length  
2672 2771: gap of unknown length  
2772 4499: contig of 1728 bp in length  
4500 4599: gap of unknown length  
4600 6309: contig of 1710 bp in length  
6310 6409: gap of unknown length  
6410 8601: contig of 2192 bp in length  
8602 8701: gap of unknown length  
8702 10415: contig of 1714 bp in length  
10416 10515: gap of unknown length  
10516 13064: contig of 2549 bp in length  
13065 13164: gap of unknown length  
13165 16156: contig of 2992 bp in length  
16157 16256: gap of unknown length  
16257 18638: contig of 2382 bp in length  
18639 21462: contig of 2724 bp in length  
21463 21562: gap of unknown length  
21563 24282: contig of 2720 bp in length  
24283 24382: gap of unknown length  
24383 28368: contig of 3986 bp in length  
28369 32469: gap of unknown length  
32469 32349: contig of 3781 bp in length  
32350 32350: gap of unknown length  
32350 36555: contig of 4206 bp in length  
36555: gap of unknown length





[illegible]

QY	224	ctacttttctatactgtgtggttaggagtcattacagtagtgactcaagaagaaacgtg	283
Db	272	AYCCNCGCNAAATAYACNGTNAARGTNGSGTNCARACNTTNCNGYAAAYWSNACNWSNG	331
QY	284	tgaagactacgtgtgcaaaatcgtcatccatcccaagtcaccaagatacaacggcagac	343
Db	332	ARTYNTNGTNCNMGNAHGTGNATHCAYGARAYTYTHAAYMGNATGWSNGAYGAY	391
QY	344	tgcctgttgtaaacgtctctcaagtcacacccctcctccatctccatccctgcctatttgc	403
Db	392	THGCNATHVNAARYTNAARTAYCCNGTNCACNTGGNSCCNTNGTNCARCCNATHGY	451
QY	404	tgcceagtgtaacaagaagcttggcaatcccccctttgttgggtgacccgagatgggaa	463
Db	452	TNCCNWSNTTYAAYTNAARCCNWSNATHGGNACNATGTGTGGGTNGTNGTGGGNY	511
QY	464	aagtaagaaagttcagatagagataccattctcctcctcaggaagcagagaccaca	523
Db	512	TNGARAARCGNARGNCAYCCNARACNCCNTAYWSNGTNCARGNYTNGCNGTNWNA	571
QY	524	ttattgaccgcaggtgttgaaacagctctacaatcccatcggtattcttcttgcagcac	583
Db	572	THGTNAAYAYGARATHGTGAAYCAYMGNATYCA-----RTTYTNYTNYTNAARAAYC	625
QY	584	tggagccagtcacaagaagagacaagatttgtgtgtgtatcactcaaaacatagaagata	643
Db	626	ARAARAARTYATHGNGAAYATGYTGTGACNWS----NWSNGARTGGGNYTNGAYA	682
QY	644	gttgcaaggtgtattctgtgagggccctctcgtgtcacattgatggttatggatccaga	703
Db	683	CNTGYCARGAYACNWSNGNWSNYSNTNGTNGYCARATGAAYARAACNTGGGTNCARA	742
QY	704	caggagtagtaagctgggattagaaatggtgtaaaatcctc-----ttcctggagtcataccaca	760
Db	743	TGGGNGTNGTNGSNTGGAATYTYGATYGGNGMNGNCARTTYCCNWSNGTNTAYACNW	802
QY	761	atgtaattactaccacaaatgattaa	788
Db	803	SNACNWSNCAITYACNCARTGGATHAA	830
RESULT	6		
LOCUS	AB038496	1850 bp	mRNA
DEFINITION	Xenopus laevis Xesp-1	mRNA for embryonic serine protease-1,	09-AUG-2000
ACCESSION	AB038496		
VERSION	AB038496.1	GI:9757697	
KEYWORDS	embryonic serine protease-1.		
SOURCE	Xenopus laevis	cDNA to mRNA.	
ORGANISM	Xenopus laevis		
REFERENCE	1 (sites)		
AUTHORS	Yamada,K., Takabatake,T. and Takeshima,K.		
TITLE	Isolation and characterization of three novel serine protease genes		
JOURNAL	Gene	252 (1-2), 209-216 (2000)	
MEDLINE	20363741		
REFERENCE	2 (bases 1 to 1850)		
AUTHORS	Yamada,K., Takeshima,K. and Takabatake,T.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-FEB-2000)	to the DDBJ/EMBL/GenBank databases.	Kazuto
	Yamada, Nagoya University, Graduate School of Human Informatics,		
	Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan		
	(E-mail:yamada@info.human.nagoya-u.ac.jp, Tel:+81-52-789-2573)		
FEATURES	Location/Qualifiers		
source	1..1850		
	/organism="Xenopus laevis"		
gene	/db_xref="taxon:8355"		
	73..1026		

CDS		/gene="Xesp-1"	
		73..1026	
		/gene="Xesp-1"	
		/note="serine protease"	
		/codon_start=1	
		/product="embryonic serine protease-1"	
		/protein_id="BAB08216.1"	
		/db_xref="GI:9757698"	
		/translation="MGKLLIYVTTLLFVSPHPSLSNITTAAPLCGSPVPSRIVGG	
		TDTRQAMPQWSELEFNHSHGGSIISDQWLLTATHCHIEHPDLPSGGGVRUGAYOLY	
		VKNHEMTVKVDIYYINSEFNPGTSGDIALKLSPIKFEYIIPICLPASVPFSS	
		ITECHWTGQGTSEVPLQYPATLQKVMVPIINRDSCKMYHINSVISETELLIQSDQ	
		ICAGYQAGKDGCGDGGPLVKIQGFYQAGIVSWGERCAAKRNPVYTFVPAYET	
		WISERSVISFRPFTSSSPSSSVLRASAILLGVSLLDHW"	
BASE COUNT	532 a	355 c	391 g 572 t
ORIGIN			
	Query Match	12.4%	Score 129.6; DB 8; Length 1850;
	Best Local Similarity	51.2%	Pred. No. 3.7e-29;
	Matches 384; Conservative	1; Mismatches 350; Indels 15; Gaps 3;	
QY	53	tgtgtgggcaaacctgtatactccagccgcgtgtgtagtggtggccaggatgctgtgcagggc	112
Db	164	TATGTGGCAGTCCAGTGTTTTCAAGTAGGATTGTTGGAGAACTGACACTAGGCAAGGAG	223
QY	113	gctggccttggcaggtcagcctacatttgaccacaactttatctatgagaggttccctcg	172
Db	224	CATGGCCTTGGCAGGTTCAGTTTAGAGTTTAATGGCAGCCATATTTGTGGGGGTCCATCA	283
QY	173	tcagtgagaggttgactgacagcagcacactgtacataccgacactggactactttt	232
Db	284	TCTCCGACCAGTGGATTCTTAACAGCGACCCACTGCATTGNACACCTGATTACCTTCIG	343
QY	233	catatactgttggctaggatgattacagtaggtgagttactcaa-----ggaaacgtg	283
Db	344	GTGTGGAGTCCGCTCTAGGGCGCTACCAGCTGTATGGAAGAACCCCTCATGAGTAGACAG	403
QY	284	tgaagtaactacgtgtccaaaatcgttcattccatccccaagtaaccaagatacaacggcagac	343
Db	404	TCAAGTTGACATTATCTACATTAACCTCAGAGTTTATTTGGCCCTGGCACAAAGTGGAGATA	463
QY	344	tcgcctgttgaaactgtcctctcaagtcacottcactctcgcattcctgcctatttgc	403
Db	464	TTGCTTTACTGAGCTGTCTAGCCCATTAATAATTCACAGATACATTCCTCCCAATATGCT	523
QY	404	tgcacagtgtaacaagaagcagttggcaattccaccctttgttgggtgacccgagtgaggaa	463
Db	524	TGCCAGCTTCTCCAGTCACTCTCTCTCTGGGACCGAATGCTGGATAACAGGATGGGGAC	583
QY	464	aagttaagaaagttcagatagagattaccattctcctcctcaggaagcagaagataccaca	523
Db	584	AGACTGGTCTGAAGTTCACATGCAATATCCAGCAACTCTTTCAGAAGGTGATGGTGGCCA	643
QY	524	ttattgaccgcaggtctgtgaacagctctacaaatcccatcggtatcttcttccagcac	583
Db	644	TCATCAACAGGGATTTCATGTGAGAAGATGTACCATATCAACTCCGTCATCAGTGAGACA	- 702
QY	584	tggagccagtcatacgaaggagaagatttggctggtgatactcaaaacatagaagata	643
Db	703	--GAAATCTGTATCCAAAGTGACCAGATCTGTGCGGGTACCAAGCAGGACAAAGGATG	760
QY	644	gttgcaaggtgtattctcgtgagggcctctcgtgtcacattgatggttatgattcaga	703
Db	761	GTGTCAGGAGACTCAGGGGGACCACTCGTCTGTGTAATAATACAAAGGGTCTTGATCAGG	820
QY	704	caggagtagtaagctgggattagaatgt---ggtaaatctcttcttcttggagttcacacca	760
Db	821	CYGGTATTGTGAGCTGGGAGAAAGATGTGCCGCTAANAACCGTCTCTGGTGTATTACTT	880
QY	761	atgtaattactaccacaaatgattaaatg	790
Db	881	TTGTTCCAGCTTATGAACATGATGATCTCTG	910

RESULT	7
LOCUS	AY005145
DEFINITION	Mus musculus testisin mRNA, complete cds.
ACCESSION	AY005145
VERSION	AY005145.1
KEYWORDS	GI:13470305
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutherii; Rodentia; Sciurognathi; Muridae; Musinae; Mus;
AUTHORS	Scarman,A.L., Hooper,J.D., Boucaut,K.J., Sit,M.L., Webb,G.C., Normyle,J.F. and Antalis,T.M.
TITLE	Organization and chromosomal localization of the murine Testisin gene encoding a serine protease temporally expressed during spermatogenesis
JOURNAL	Eur. J. Biochem. 268 (5), 1250-1258 (2001)
MEDLINE	21153229
PUBMED	11231276
REFERENCE	2 (bases 1 to 1055)
AUTHORS	Scarman,A.L., Hooper,J.D., Webb,G.C., Normyle,J.F., Sit,M.L. and Antalis,T.M.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2000) Cellular Oncology Laboratory, Queensland Institute of Medical Research, 300 Herston Road, Brisbane, Qld 4006, Australia
FEATURES	Location/Qualifiers
source	1..1055
	/organism="Mus musculus"
	/strain="129/Sv"
	/db_xref="taxon:10090"
	/chromosome="17"
	/map="A2-B"
CDS	1..975
	/note="putative serine protease"
	/codon_start=1
	/product="testisin"
	/protein_id="AA02255.1"
	/db_xref="GI:13470306"
	/translation="MGARKTLVPLVVVATAMALQSYLQVDPEKPELOEPDILLSG PCGHRTPRIVGGDDAEILGVLPWQGLRWGNHLCATLLNRNWLTAHCFOKDNI PDMTVQFGLTSRPSLNLAQYSNRYQIEDIFLSPKYSEQYPNDIALKLSPVPYY NFPTQCLLNSITYKFENRTDCWVTGWGAIGDESLSPNTLOEVQAIIINSMCMHY KKPDFRTNWGMVCAGTPEGGRKADCFGSPPGLACDQTVMFTQCVSVWGICGCRPN RGCVTNISHNYNQISTMRNGLRPRDPVPLLFLTLAWSSLLRPA"
BASE COUNT	228 A 292 G 253 C
ORIGIN	
Query Match	12.4%; Score 129.4; DB 94; Length 1055;
Best Local Similarity	52.9%; Pred.No. 4e-29;
Matches 392; Conservative	1; Mismatches 312; Indels 36; Gaps 4;
QY	78 ccgcgttgtagtggccagatgctgctgcaggcgctgcccttgccagtgcaggcaccaccca 137 
Dd	159 CCGTATTAGTGTTGGCGCATGCTGACTGTGCCGTGCCTGCCGTCGCAAGGAGCCTCG 218 
QY	138 ctttgaccacaacttatctatgatgaggttcctctogtcaagtgaaggttgatatgacagc 197 
Dd	219 TGTATGGGGCAACACATTATGTGGCGCAACCTTGCTCAACCCGCCGTGGGTGCTTACAGC 278 
QY	198 agcacatgcatacaaccgacctggactatttctaatacttgtgtgcttagcatgatgat 257 
Dd	279 TGCCCCACTGCTTCCAAGAAGATAACGATCCTTTTGACTGGACAGTCCAGTTTGGTAGACT 338 
QY	258 tacagtaggtgactccaaggaaacgtgtgaagtacta-----cggtccaaaatacgtca 310 
Dd	339 GACTTCCAGGCCATCTCTCTGGAACTACAGGCCGTATTCACACCGTTACCAAATAGAAGA 398 
QY	311 tccatcccagaagtaccaagatacaaacggca-----gaortcgccctgttgaaact 359 

[illegible]

[illegible]

```

QY 568 atcttctgacagcagctgagccagtcacatcaaggaagacacaaagtattgtgctggtgatact 627
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 637 TTCCGCAAG-----GACATCTTTGGAGACATGTTTGTGCTGGCAATGCC 681
QY 628 caaacatgaagatagttgcaagggtgattctgaggggccctctgctgctcacattgat 687
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 682 CAAGCGGGAAGGATGCCCTTCCTGGTGACTCAGGTGGACCCCTTGCCCTGTAAACAAGAAAT 741
QY 688 ggtgatgacacagacagagtagtaagctgggagtagaattggttaaatc---tctt 744
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 742 GCACGTGTATCAGATTGGAGTCTGAGCTGGGAGTGGGCTGTGGTCGGCCCAATCGG 801
QY 745 cctggagctacaccaatgaattactactaccacaaatggat 785
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 802 CCCGGTGTACACCAATATACGCCACCACCTTTGAGTGGAT 842

RESULT 10
E21866
LOCUS E21866 1082 bp DNA PAT 07-FEB-2001
DEFINITION Novel acidophil serine protease.
ACCESSION E21866
VERSION E21866.1 GI:13023737
KEYWORDS JP 199032768-A/2.
SOURCE unidentified.
ORGANISM
REFERENCE
AUTHORS Hiroshi, K.M.I.
TITLE Novel acidophil serine protease
JOURNAL Patent: JP 199032768-A 2 09-FEB-1999;
COMMENT ONO PHARMACEUT CO LTD
OS Unidentified
PN JP 199032768-A/2
PD 09-FEB-1999
PF 16-JUL-1997 JP 1997191319
PR
PI HIROSHI KIDO,MASAHIRO INOUE
PC C12N15/09,A61K38/55,A61K39/395,A61K48/00,C07K7/00,
PC C07K16/40,
PC C12N9/64,C12N15/00,A61K37/64,A61K37/64
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..1082
FEATURES
source Location/Qualifiers
1..1082
/organism="Unidentified".
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 216 a 318 c 308 g 240 t
ORIGIN

Query Match 11.4%; Score 118.6; DB 10; Length 1082;
Best Local Similarity 51.9%; Pred. No. 9.8e-26;
Matches 395; Conservative 1; Mismatches 320; Indels 45; Gaps 4;

QY 55 tttgggaacctgtatactccagccgctgttagtggccagtagctgctcagggcgc 114
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 106 TCGCGCGACGGGTGATCAGCTGCGCATCGTGGTGGAGAGGACGCCGAATCGGGCGT 165
QY 115 tggccttgaggtcagctcaccttggaccacactttatctatctgaggttccctcgtc 174
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 166 TGGCGGTGGCAGGGAGCGCTGCGCTGTGGGATTCCACGTATGCGGAGTGACCTGTCTC 225
QY 175 agtgaaggttgatactgacagcagcagcagcagcagcagcagcagcagcagcagcagc 234
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 226 AGCCACCGCTGGGCACCTACGCGCGGCGACTGCTTTGAAACCTATAGTACCTTAGTGAT 285
QY 235 tatactgtgtgctaggtgattacagtaggtgactcaaggaacgtgtgaag----- 288
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 286 CCTCCGGGTGGATGCTCCAGTTTGGCCAGCTGACTTCCATGCCATCCTTCTGGAGCCTG 345

```

```

QY 289 -----tactacgtgtcctcaaatc---gtcatccatcccaagtaccaa 327
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 346 CAGGCCTACTACACCGTTACTTCTGATCGAATATCTATCTGAGCCCTCGCTACCTGGGG 405
QY 328 gatacaacggcagacortcgccctgttgtaactgtcctcaagtcaaccttccactttgtgc 387
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 406 AATTCACCCCTATGACATTGCCCTGGTGAAGCTGCTCTGCACCTGTACCTTACACTAAACAC 465
QY 388 atctgcctattgttggcccgagtcacaaagcagttggcaattccacccctttttgtgg 447
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 466 ATCCAGCCCATCTGCTCCAGGCCCTCCACATTGAGTTTGAGAACCGGACAGACTGCTGG 525
QY 448 gtgacgagtaggggaaagttaagaaagttcagatagagattaccattctgccccttcag 507
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 526 GTGACTGGCTGGGGGTACATCAAGAGGATGAGGCACTGCCATCTCCCCACACCCCTCCAG 585
QY 508 gaagcagaagtagccattattgacgcgcccaggtgttgaaacagctctcctgctgacacattg 567
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 586 GAAGTTTCAGTGGCCATCATAAACAACCTATATGTGCAACCACTCTTCTCAAGTACAGT 645
QY 568 atctcttgcagcagcagtgagccagtcacaaaggaagacaaagtattgtgctggtgatact 627
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 646 TTCCGCAAG-----GACATCTTTGGAGACATGTTTGTGCTGGCAATGCC 690
QY 628 caaacatgaagtagttgcaagggtgattctgaggggccctctcctgctgacacattgat 687
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 691 CAAGCGGGAAGGATGCCCTGCTTCGTGACTCAGGTGGAGCCCTTGCCCTGTACACAGAAT 750
QY 688 ggtgatgacacagcagagtagtaagctgggagtagaattggttaaatc---tctt 744
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 751 GGACTGTGTTATCAGATTGGAGTGGTGGAGTGGGAGTGGCTGTGTCGGCCCAATCGG 810
QY 745 cctggagctacaccaatgaattactactaccacaaatggat 785
  ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 811 CCCGGTGTCTACCAATATACGCCACCACCTTTGAGTGGAT 851

RESULT 11
E21867
LOCUS E21867 1082 bp DNA PAT 07-FEB-2001
DEFINITION Novel acidophil serine protease.
ACCESSION E21867
VERSION E21867.1 GI:13023738
KEYWORDS JP 199032768-A/3.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Hiroshi, K.M.I.
TITLE Novel acidophil serine protease
JOURNAL Patent: JP 199032768-A 3 09-FEB-1999;
COMMENT ONO PHARMACEUT CO LTD
OS Homo Sapiens
PN JP 199032768-A/3
PD 09-FEB-1999
PF 16-JUL-1997 JP 1997191319
PR HIROSHI KIDO,MASAHIRO INOUE
PC C12N15/09,A61K38/55,A61K39/395,A61K48/00,C07K7/00,
PC C07K16/40,
PC C12N9/64,C12N15/00,A61K37/64,A61K37/64
CC Topology: Linear;
FH Key Location/Qualifiers
FT CDS 10..951
FT sig peptide 10..87
FT mat peptide 133..951.
FEATURES
source Location/Qualifiers
1..1082
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 216 a 318 c 308 g 240 t

```



QY	388	atcctgcctattgttgcctccagtgctcacaaagccagttggcaattccacccttttgttg	447	GEDDELGNPHQGSLSRLWDSHVCVSLLSHRWALTAAHCFETYSDLSDFSGWVQGG
Db	466	ATCCAGCCCATCTGCCTCCAGCCCTCCACATTTTGTAGTTTGAGAACCGGACAGACTGCTGG	525	LTSMPSPWSLOAYITRYFVNIYSPRLGNSPDIALVKLSAPVTHVTHHIIQICLQ
QY	448	gtaccggatggggaagattgaagaaagttcagatagattaccatttgccttcag	507	STFFENRTDCWTWGYIKEDALPSPHTLQVQVAIINNSMNLFLKYSFRKDI
Db	526	GTACCTGGCTGGGGGTACATCAAGAGGATGAGGCACATGCCCATCTCCCAACCCCTCCAG	585	GDMYCAGNAOGKDFGDSGGPLACNKNGLWYQIGVSVSGVGRNRPVGVYTNISH
QY	508	gaagcagaagatccattattgaccgccaggtgttggaacagctctacaatccccatcggt	567	HFIEWIKLMAQSGMSQDPDPSWPLLPFLFWALPLLGPV
Db	586	GAAGTTTCAGGTGCGCCATCAAAACAACTCTATGTGCAACACCCTCTTCTCAAGTACAGT	645	1058
QY	568	atcttttccagcactggagccagtcacatcgaaggaagacaagattgtgtggtgatact	627	/note="13 a nucleotides"
Db	646	TTCCGCAAG-----GACATCTTTTGGAGACATGTTTGTGCTGGCAATGCC	690	BASE COUNT 208 a 317 c 305 g 241 t
QY	628	caaaacatgaagagatgttgcaagggtgattctgagagggcctctgtgtgcacattgat	687	ORIGIN
Db	691	CAAGCGGGGAAGGATGCCCTGCTTCGGTGACTCAGGTGACCCCTTTGGCCCTGTAAACAAGAT	750	Query Match 11.2%; Score 117; DB 85; Length 1071;
QY	688	ggtgtgatccagcagagtagtaagactgggattagaaatgtggttaaatc---tctt	744	Best Local Similarity 51.8%; Pred. No. 3.le-25;
Db	751	GGACTGTGTATCAGATTGGAGTGTGAGCTGGGAGTGGGCTGTGGTGGGCCCAATCGG	810	Matches 394; Conservative 1; Mismatches 321; Indels 45; Gaps 4;
QY	745	cctggagctcacaccaatgaattactactacacaaaatgat	785	
Db	811	CCCGGTGCTACACCAATATCAGCCACCACTTTGAGTGGAT	851	
RESULT 13				
AB031330				
LOCUS				
DEFINITION				
Homo sapiens esp-1 mRNA for eosinophil serine protease, complete cds				
AB031330 1071 bp mRNA PRI 15-JAN-2000				
Homo sapiens esp-1 mRNA for eosinophil serine protease, complete cds				
AB031330				
AB031330.1 GI:5777331				
eosinophil serine protease.				
Homo sapiens cell_line:HeLa S3 cDNA to mRNA, clone_lib:HeLa cDNA				
Lambda Triplex.				
Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
1 (sites)				
Inoue,M., Isobe,M., Itoyama,T. and Kido,H.				
Structural analysis of esp-1 gene				
Biochem. Biophys. Res. Commun. 266 (2), 564-568 (1999)				
20068805				
2 (bases 1 to 1071)				
Inoue,M., Isobe,M. and Kido,H.				
Direct Submission				
Submitted (20-AUG-1999) to the DDBJ/EMBL/GenBank databases.				
Masahito Inoue, Institute for Enzyme Research, Division of Enzyme Chemistry; 3 Kuramoto-cho, Tokushima, Tokushima 770-8503, Japan				
(E-mail:inoue@ier.tokushima-u.ac.jp, Tel:81-886-33-7424, Fax:81-886-33-7425)				
FEATURES				
Location/Qualifiers				
1..1071				
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
/cell_line="HeLa S3"				
/chromosome="16"				
/clone_lib="HeLa cDNA Lambda Triplex"				
/map="l6p13.3"				
4..948				
/gene="esp-1"				
4..948				
/gene="esp-1"				
/codon_start=1				
/product="eosinophil serine protease"				
/protein_id="BA083521.1"				
/db_xref="GI:5777332"				
/translation="MGARGALLALLARLARLRKPESQEAAPLSGGPCGRRVITSRVIG				
gene				
CDS				
RESULT 14				
AF058300				
LOCUS				
AF058300 1077 bp mRNA PRI 01-JUL-1999				
DEFINITION Homo sapiens testisin (TEST1) mRNA, complete cds.				

```

VERSION      AF058300.1  GI:5305322
KEYWORDS     human.
SOURCE       Homo sapiens
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 1077)
AUTHORS     Hooper J.D., Dickinson J.L. and Antalis T.M.
TITLE       Direct Submission
JOURNAL     Submitted (08-APR-1998) Queensland Institute of Medical Research,
              300 Herston Road, Brisbane, QLD 4006, Australia
FEATURES     Location/Qualifiers
              1..1077
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="16"
               /map_16p13.3
               /cell_line="HeLa S2"
              1..1077
               /gene="TEST1"
              19..963
               /gene="TEST1"
               /note="serine proteinase"
               /codon_start=1
               /product="testisin"
               /protein_id="AAD41588.1"
               /db_xref="GI:5305323"
               /translation="MGARGALLALLARAGLRKPESOEAPLPGCGRRVITSRIVG
              GDALGRPWGSLRLWDSHVCGVSLSHRWALTAHCFEYSLSDPSCHWVQFG
              LTMSPFWSLQAYTRYFVSNLYLSPRYLGNPSYDIALVKLSAPVYTHKIQICLQA
              STFEENRTDVTWGTYKEDALPSPHTLQEVQVAIINSMCNHLFLKYSFRKDI
              GDMVCAAGNAQGCACFCFDSGGPLACNKLNGLYQIVSVGVCGRPNRPYVYTNISH
              HFEWIKLMAQSGMSQPDPSWPLFLFPLLWALPLGVP"
              polyA_signal
              202 a 320 c 314 g 241 t
              /gene="TEST1"
BASE COUNT   202 a 320 c 314 g 241 t
ORIGIN
Query Match      11.2%; Score 117; DB 88; Length 1077;
Best Local Similarity 51.8%; Pred. No. 3.1e-25;
Matches 394; Conservative 1; Mismatches 321; Indels 45; Gaps 4;

Qy 55 tgtgggcaacctgtatactccagccgctgttaggtggcaggatgtctgtcagggcgc 114
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 115 TGGCGCGACGGGTATCATCGTCCGCGATCGTGGTGGAGAGAGCGCGAATCGGCGT 174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 115 tggccttggcaggtcagcctacactttgacacaaactttatctatggaggttcctcgtc 174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 175 TGGCGGTGGCAGGGAGCGCTCGCGCTGTGGGATTCCACAGTATGCGGAGTGAGCCTGCTC 234
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 175 agtgagaggttgatactgacagcagcactgcatacaaccgacctggactatttca 234
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 235 AGCCACCGCTGGGCACATCAGCGGGCGCATCTGCTTTGAAACCTATAGTACCTTAGTGAT 294
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 235 tatactgtggctaggatcgattacagtagtgagtcacaaaggaaacgtgtgaag----- 288
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 295 CCCTCCGGGTGGATGCTTGGCCAGTGCTACTTCCATGCCATCCTTCTGGAGCGCTG 354
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 289 -----tactacgtgtcaaaaatc---gtcatccatcccaagtaaccaa 327
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 355 CAGCGCTACTACACCCGTTACTTCGTATCGAATATCATCTGAGCCCTCGCTACCTGGGG 414
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 328 gatacaacggcagcrtcgcttgttgaacttcctctcaagtcacacttcactctcgc 387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 415 AATTCACCTATGACATGTCCTGGTGAAGCTGTCTGCACCTGTACCTACATAACAC 474
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 388 atctgtcctattgtgtccaggtgtcacaagcagttggaattccaccctttgttgg 447
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 475 ATCCAGCCCATCTCTCTCCAGGCGCTCCACATTTTGAGTTTGAGAACCGGACAGAGCTGCTGG 534
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 448 gtgaccggatgggaaagttaagaaagttcagatgagatgattaccatttgccttcag 507
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

Db 535 GTGACTGGCTGGGGGTACATCAAGAGGATGAGGCACCTGCCATCTCCACACACCCTCCAG 594
Qy 508 gaagcagaagtagccattattgacgcgcaggcttgtgaacagctctacaatccccatcggt 567
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 595 GAAGTTCAGGTCCGCATCATAAACAACTCTATGTGCAACCACTCTTCTCAAGTACAGT 654
Qy 568 atcttctgtccagcactggagccagctcatcaagggaagacaagatttgtgtgtgatact 627
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 655 TTCCGCAAG-----GACATCTTTGGAGACATGTTTGTGCTGGCAATGCC 699
Qy 628 caaacatgaagatagttgcaagggtgattctggaggcctctgtcggtgcacattgat 687
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 700 CAAGCGGGAAGATGCCTCTCGGTGACTCAGGTGGAGCCCTTGGCCCTGTAAACAAGAT 759
Qy 688 ggtgtatggtccagacagagtagtagtaactggggattagaatgtggtgtaaatc---tcct 744
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 760 GGACTGTGCTATCAGATTGGAGTCGTGAGCTGGGAGTGGGCTGTGGTGGGCCAATCGG 819
Qy 745 cctggagctcaccaaatgtatctactactacacaaaatggat 785
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 820 CCGGTGTCTACACCAATATCAGCCACCACCTTTGAGTGGAT 860
```

```

RESULT 15
HSA306593      1013 bp      mRNA      PRI      31-MAR-2001
LOCUS          Homo sapiens mRNA for marapsin (MPN gene).
DEFINITION     AJ306593
ACCESSION      AJ306593.1  GI:13516325
VERSION        marapsin; MPN gene.
KEYWORDS       human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1013)
AUTHORS        Fortunato M., Dando P.M., Rawlings N.D. and Barrett A.J.
TITLE          Cloning, sequencing and expression of marapsin, a human serine
               proteinase
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 1013)
AUTHORS        Fortunato M.
TITLE          Direct Submission
JOURNAL        Submitted (29-MAR-2001) Fortunato M., MRC Molecular Enzymology
               Laboratory, Babraham Institute, Babraham Hall, Babraham, Cambridge,
               CB2 4AT, UNITED KINGDOM
COMMENT        Related entry AW170323.
FEATURES       Location/Qualifiers
               1..1013
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE 2698030"
                20..892
                /gene="MPN"
                20..892
                /gene="MPN"
                /function="putative trypsin-like activity"
                /codon_start=1
                /product="marapsin"
                /protein_id="CAC35467.1"
                /db_xref="GI:13516326"
                /translation="MRRPAAPLPLLLLCFQSRAKAATACGRPRMLRMVGGQDTQEG
               ENPQVSIQRNSHFQCGSLLAEOWVLTAAHCPRNTSETSLYOVLLGAROLVQPGPHA
               MYARQVESNPDIQGTASSADVALVELEAPVFTNYILPVLCPDPSVIFETCMQWV
               TGMSPSEEDLLPEPRILQKLAVPIIDTPKCNLLYSKDFEFGQPKTIKNMLCAGFE
               ECKDKACKDGGPLVCLVQSQWLVISWEGECARQNRPGVYIRVTAHHNWIHRII
               PKLQFQPARLGGOK"
BASE COUNT     192 a 334 c 321 g 166 t
ORIGIN
Query Match      11.1%; Score 115.8; DB 93; Length 1013;
Best Local Similarity 50.6%; Pred. No. 7.4e-25;
Matches 388; Conservative 1; Mismatches 363; Indels 15; Gaps 4;
```



Qy	50	cag	tgtgtg	gcac	ctgtata	ctccag	ccgcgtgt	gtagt	gtgccc	aggtat	gtctgt	cgag	109
Db	90	cag	ctt	tggtgc	gccc	cagga	tgtga	accgaat	gtgtg	gcggc	gacac	gcag	149
Qy	110	ggc	gtgc	cccttg	gcag	ctcag	ctacac	ctttg	aacca	aaactt	tatctat	gaggtt	169
Db	150	gcg	agt	ggccct	tgca	agtcag	gatcc	ggcg	aacggaa	gccact	ttctgcg	ggggcag	209
Qy	170	tcg	tac	gtagag	gttgata	ctact	gacag	cagca	cactgt	catata	caacgc	actgg	229
Db	210	tc	atgc	ggagc	agctgg	gtctg	acg	gcctgc	gcgac	tgtctcc	ccacac	acctctg	269
Qy	230	ttt	catata	ctgtgt	gctg	atgattac	agtag	gttgact	oaagg	aaacgt	gtga	agt	289
Db	270	cc	ctgttacc	aggtct	ctgtgc	ggggc	aaggc	agctag	tgtgac	ggcgac	ccacac	cgctat	329
Qy	290	a	---ct	agtg	tccaaa	atcg	tccat	cccca	agtacca	agata	caa	g	340
Db	330	atg	ccccgg	gtgag	caggtg	gagac	aaccccc	ctgtacc	agggc	acggc	ctccag	cg	389
Qy	341	ac	rtgc	ctt	gttga	aactgt	ctctc	aa	gtccact	cttctgc	acttgc	tatt	400
Db	390	ac	gtgg	ccct	ggtag	ctgga	ggc	accag	tgcctt	cacc	aat	tac	449
Qy	401	g	cttgc	ccag	tgtca	aaa	gcag	tgtgg	caatt	ccacc	ctttgt	gtg	460
Db	450	gc	cttgc	tga	cccc	ctcg	tgatct	ttt	gagac	gggc	atga	act	509
Qy	461	g	aaa	gtta	agg	aaagt	tcag	a	tag	attac	attc	gtcc	520
Db	510	gc	ac	cccc	agtga	ggaa	gac	ctc	ctgc	ggat	tc	ctgc	569
Qy	521	cc	att	tgacc	gcag	gctgtg	aac	agct	ctctaca	atcccc	atcg	tat	580
Db	570	cc	at	cat	cgac	acac	ccaa	agtga	ccact	gct	ctaca	--	626
Qy	581	c	act	tgag	ccag	tcat	ca	agg	aagaca	agatt	gtg	ctg	640
Db	627	ac	ca	accc	caaa	aac	ctac	aga	aat	gac	at	gtg	686
Qy	641	a	tag	ttg	ca	aggt	gtatt	ct	tgag	gccc	ctct	gt	700

**This Page Blank (uspto)**

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	585	31.4	389	13	Q9Pvx7	Q9pvx7 xenopus lae	
2	554.5	29.8	317	13	Q9DGR3	Q9dgr3 xenopus lae	
3	541.5	29.1	305	11	Q9JHU7	Q9juj7 mus musculus	
4	526	28.2	314	4	Q9Y6M0	Q9ym60 homo sapien	
5	523	28.1	312	4	Q9NS34	Q9ns34 homo sapien	
6	522	28.0	339	11	Q9ESD1	Q9esd1 mus musculus	
7	510.5	27.4	310	11	Q9QY29	Q9qyz9 mus musculus	
8	504	27.1	342	11	Q9ER01	Q9er01 rattus norv	
9	503	27.0	342	11	Q9ES87	Q9es87 rattus norv	
10	493.5	26.5	311	11	Q9QUL7	Q9qu17 mus musculus	
11	492	26.4	297	11	Q9B8781	Q9b8781 rattus norv	
12	491.5	26.4	321	4	Q9UBB2	Q9ubb2 homo sapien	
13	490.5	26.3	306	11	Q9ER10	Q9er10 mus musculus	
14	490.5	26.3	321	4	Q9NR08	Q9nrq8 homo sapien	
15	488.5	26.2	317	4	Q9GZM4	Q9gzr4 homo sapien	
16	477.5	25.6	321	4	Q9NRR2	Q9nrr2 homo sapien	
17	477	25.6	300	4	Q9P2V6	Q9p2v6 homo sapien	
18	472	25.3	275	4	Q9UQ11	Q9uq11 homo sapien	
19	468	25.1	273	6	Q9XSM2	Q9xsm2 ovis aries	



[illegible]

DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE CHANNEL ACTIVATING PROTEASE 1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Vuagniaux G., Vallet V., Fowler-Jaeger N., Bens M., Farman N.,  
RA Courtiol-Couty N., Vandewalle A., Rossier B.C., Hummler E.;  
RT "Activation of the anilorhine-sensitive sodium channel by the mouse  
RT serine protease mCAP1 expressed in a principal kidney cell line."  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF18613; AAG17054.1; -.  
KW Protease.  
SQ SEQUENCE 339 AA; 36234 MW; B64D187D0F4CE62 CRC64;

Query Match 28.0%; Score 522; DB 11; Length 339;  
Best Local Similarity 40.8%; Pred. No. 7.3e-42;  
Matches 111; Conservative 44; Mismatches 99; Indels 18; Gaps 6;

QY 7 AFTLLLLGI-----SVCQGPVYSSRVVGGQDAAGRPWQVSLHFDHFIYGG 55  
DB 14 AVTLLLLGLLQSGIRADGTEASCG-AVIOPTITGGGSAKPGQWQVSIYDGNHVCVG 72  
QY 56 SLVSERLILTAACHIOPTWTFYSYTWLGSITVGD-SRKRKYVYKVIHPKYQD--TT 112  
DB 73 SLVSNKWVSAACHCFPREHSREAYEVLGAHQDLSYSDNTVVTVAQIITHSSYREESQ 132  
QY 113 ADVALLKLSQVTFTSAILPICLPSTVKQLAIPFCWVTGKVKKSSDRDYHSAEQAE 172  
DB 133 GDIAFIRLSFPVTSYRIRPICLPAANASPNGLHCTVTGNGHVAPSVLSQTRPRLQGLE 192  
QY 173 VPIIDRACQOLYNPIGIFLPALEPIKEDKICAGDTQNMKDSCKDGGSLCHIDGVW 232  
DB 193 VPLISRTCSCLNINAV--PEPHTIQDMLCAGYKVGKDGACQDGGSLPCPMEGIW 250  
QY 233 IQTGVSWSGLECG-KSLPGVYTNVYKQWIN 263  
DB 251 YLAGIVSGWDACGAPNRPVGVYTLTSTVASHIH 282

RESULT 7  
Q9QY29 PRELIMINARY; PRT; 310 AA.  
AC Q9QY29  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE DISTAL INTESTINAL SERINE PROTEASE.  
GN DISP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Shaw-Smith C.J., Coffey A.J., Leversha M., Freeman T.C., Bentley D.R.,  
RA Walters J.R.;  
RT "Characterization of a novel murine intestinal serine protease,  
RT DISP.";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).  
DR EMBL; AJ243866; CAB56465.1; -.  
DR HSSP; P20231; 1A0.  
DR InterPro; IPR000038; -.  
DR InterPro; IPR001254; -.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.

DR ProDom; PD002565; -; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
DR SMART; SM00020; Tryp\_SPC; 1.  
KW Protease.  
SQ SEQUENCE 310 AA; 33701 MW; F828EC7F6D25303F CRC64;

Query Match 27.4%; Score 510.5; DB 11; Length 310;  
Best Local Similarity 38.8%; Pred. No. 8.3e-41;  
Matches 120; Conservative 45; Mismatches 113; Indels 31; Gaps 11;

QY 4 AGCAFTLLLLGI-----SVCQGPVYSSRVVGGQDAAGRPWQVSLHFDHFIYGG 54  
DB 5 ARCIFLLLLQILTRARGDILPSVCGHSRDAGKIVGGQDALEGQMPQVSLWITEDGHICG 64  
QY 55 GSIVSERLILTAACHIOPTWTFYSYTWLGSITVGDSSRKR-VKYYVSKVIHPKY---QD 110  
DB 65 GSIIHEVWVLTAACHFRSLNPSFYHVKGGLLSLLEPHSTILVAVRNIFVHTYLWADA 124  
QY 111 TTADVALLKLSQVTFTSAILPICLPSTVKQLAIPFCWVTGKVKKSSDRDYHSAEQE 170  
DB 125 SSGDIALVQLDTPLR-PSQFTPVCLPAAQTPLPGTVCWVTGWGATQE---RDMASVLOE 180  
QY 171 AEVPIIDRACQOLYNPIGIFLPALEPIKEDKICAGDTQNMKDSCKDGGSLCHIDG 230  
DB 181 LAVPLLDSEDCRMWYHTQGSLSG-ERIIQSDMLCAGYVEGHIDSCQDGGPLVCSINS 239  
QY 231 VMTQGVWSWGLECGKSL-PGVYTNVYKQWINATISRRANLDFSDF-----LF 279  
DB 240 SMTQGVGTSWGIICARPYRPGVYTRVTVYVDWVQIRLAE-NHSDAYGYHSSASAYQMLL 298  
QY 280 PIVLLSLAL 288  
DB 299 P-VLLAVAL 306

RESULT 8  
Q9ER01 PRELIMINARY; PRT; 342 AA.  
AC Q9ER01  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE PROSTASIN PRECURSOR.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=KIDNEY;  
RA Adachi M., Kitamura K., Miyoshi T., Tomita K.;  
RT "Rat serine protease.";  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB017638; BAB20281.1; -.  
SQ SEQUENCE 342 AA; 36843 MW; 5EDIAF05D9213B98 CRC64;

Query Match 27.1%; Score 504; DB 11; Length 342;  
Best Local Similarity 40.4%; Pred. No. 3.9e-40;  
Matches 110; Conservative 43; Mismatches 101; Indels 18; Gaps 6;

QY 7 AFTLLLLGI-----SVCQGPVYSSRVVGGQDAAGRPWQVSLHFDHFIYGG 55  
DB 14 ALFTLLIGLLQIRGADGTEASCG-AVIOPTITGGGSAKPGQWQVSIYNGVHVCVG 72  
QY 56 SLVSERLILTAACHIOPTWTFYSYTWLGSITVGD-SRKRKYVYKVIHPKYQD--TT 112  
DB 73 SLVSNQWVSAACHCFPREHSKEEYEVKLGAHQDLSFSDIVVHTVAQIISHSYREESQ 132  
QY 113 ADVALLKLSQVTFTSAILPICLPSTVKQLAIPFCWVTGKVKKSSDRDYHSAEQAE 172  
DB 113 ADVALLKLSQVTFTSAILPICLPSTVKQLAIPFCWVTGKVKKSSDRDYHSAEQAE 172



```
DR HSP: P00761; LEPT.  
DR MEROPS: S01.252; -.  
DR InterPro: IPR001254; -.  
DR InterPro: IPR001314; -.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.  
DR PROSITE: PS00135; TRYPSIN_SBR; 1.  
DR SMART: SM00020; Tryp_SPC; 1.  
KW Hydrolyase; Protease; Serine protease; Signal.  
FT NON_TER 1  
FT SIGNAL <1 23  
FT CHAIN 24 297  
FT CHAIN 24 297  
SQ SEQUENCE 297 AA; 32086 MW; 6DA25C6633D6AB55 CRC64;  
  
Query Match 26.4%; Score 492; DB 11; Length 297;  
Best Local Similarity 35.8%; Pred. No. 4.6e-39;  
Matches 100; Conservative 58; Mismatches 101; Indels 20; Gaps 6;  
  
QY 3 PACCAFTLLI-----LLGISVCGOPVYSRVVGGQDAAGRWPMQVSLHFDHFIY 53  
DB 6 PPGTUTILLPSVTVAANIRGSPDCGKPOQLRVVGGDSADAPWIVSILKNGSHHC 65  
  
QY 54 GGSILVSRLLITAAHCIOPTWTTFS-YTVWLGSIYVDSRRK-VKYVVSIVIPKRY--- 108  
DB 66 AGSLTNRWVSAAHCFSSNMWKPSPVSVLLGAWKLGNGPGRSKQVGIASVLPHPYSRK 125  
  
QY 109 QDITADVALLKLSQVFTTSAILPICLPSTVKQLAIPPCFWCTGKVKKSSDRDYHSAL 168  
DB 126 EGHADIALVRLERPIQFSERILPICLPDSSVHLPPNTNCIAGWSIQDGVPLPRQTL 185  
  
QY 169 QEAEVPIIDROACEQLY-NGIFLPALEPIVKEDKICAGDTQNMKDSCKDGGPLSCH 228  
DB 186 QKLKVIIDPELCKSLY-----WRGACQEAITEDMCAVLEGKRGKRDACLDGSGPLMCQV 240  
  
QY 229 DGVWITQGVVSWGLECG-KSLPGVYTNVYVYQKWINATI 266  
DB 241 DDHLLTGIISWGEGCAERNRPGVYTSLLAHRPWVQRIV 279  
  
RESULT 12  
Q9UBB2 ID Q9UBB2 PRELIMINARY; PRT; 321 AA.  
AC Q9UBB2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE TRANSMEMBRANE TRYPTASE.  
GN TMT.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99452974; PubMed=10521469;  
RA Wong G.W., Tang Y., Peyfant E., Sali A., Li L., Li Y., Huang C.,  
RA Friend D.S., Krillis S.A., Stevens R.L.;  
RT "Identification of a New Member of the Trypsin Family of Mouse and  
RT Human Mast Cell Proteases Which Possesses a Novel COOH-terminal  
RT Hydrophobic Extension."  
PL J. Biol. Chem. 274:30784-30793(1999).  
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.  
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).  
DR EMBL; AF175759; AAF03697.1; -.  
DR EMBL; AF175522; AAF03695.1; -.  
DR HSP: P00763; IDPO.  
DR InterPro: IPR001254; -.  
DR InterPro: IPR001314; -.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
```

```
DR SMART: SM00020; Tryp_SPC; 1.  
KW Hydrolyase; Serine protease; Transmembrane.  
SQ SEQUENCE 321 AA; 33827 MW; FFF7B06E3C4A962D CRC64;  
  
Query Match 26.4%; Score 491.5; DB 4; Length 321;  
Best Local Similarity 36.9%; Pred. No. 5.7e-39;  
Matches 116; Conservative 49; Mismatches 114; Indels 35; Gaps 12;  
  
QY 1 MGPAGCAFTLLILL-GISV-----CGQPVYS---SRVVGQDAAGRWPMQVSLHFDHN 50  
DB 1 MALGACGLILLLLAVPGVSLRTLOPGCGRPQVSDAGRIYVGGHAAPAGANPQASLRMRM 60  
  
QY 51 FIYGGSLVSRLLITAAHCIOPTWTTFSYTVWLG--SITVGDSSRKRYKYVSVKIVH--P 106  
DB 61 HVGCGSLLSPQWVLTAAHCFSGSLNSDSYQVHLGELEITLSPHFSTVR----QIILHSSP 116  
  
QY 107 KYQ-DITADVALLKLSQVFTTSAILPICLPSTVKQLAIPPCFWCTGKVKKSSDRDYH 165  
DB 117 SGQPGTSGDIALVELSPVPTLSSRLIPVCLPEASDDFCPGIRCWVGTGWTGEGEPLPPP 176  
  
QY 166 SALQEAEPVPIIDROACEQLY-NGIFLPALEPIVKEDKICAGDTQNMKDSCKDGGPL 224  
DB 177 YSREVKVSVVDTCRRDYPGPG-----SILQPDMLCA---RPGDACQDSDSGPL 226  
  
QY 225 SCHIDGWIQTGVVSWGLECG-KSLPGVYTNVYVYQKWINATISRANNLDFSLFPIV- 282  
DB 227 VCQVNGAWQAGIVSWGEGCGRPNRPGVYTRVPAYVNWIRRHITASGSGESGYPLPLLA 286  
  
QY 283 ---LLSLALLCPSC 293  
DB 287 GLFLPLGLFLLLVSC 300  
  
RESULT 13  
Q9ER10 ID Q9ER10 PRELIMINARY; PRT; 306 AA.  
AC Q9ER10;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE BRAIN SPECIFIC SERINE PROTEASE-4 TYPE1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRIN;  
RA Mitsui S., Okui A., Kominami K., Yamaguchi N.;  
RT "Cloning and characterization of a novel serine protease, mBSP-4,";  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB010778; BAB20262.1; -.  
KW Protease.  
SQ SEQUENCE 306 AA; 33262 MW; FBBF03C0C285E7E8 CRC64;  
  
Query Match 26.3%; Score 490.5; DB 11; Length 306;  
Best Local Similarity 34.4%; Pred. No. 6.7e-39;  
Matches 97; Conservative 61; Mismatches 99; Indels 25; Gaps 5;  
  
QY 5 GCAFTLLILLISV-----CGQPVYSRVVGGQDAAGRWPMQVSLHFDH 49  
DB 12 GDQFSILILLVLLTSTAPISAATIRVSPDCGKPOQLNRIVGGEDSMDAQWPIVSIKNG 71  
  
QY 50 NFYIYGGSLVSRLLITAAHCIOPTWTTFS-YTVWLGSIYVDSRRK-VKYVVSIVIPK 107  
DB 72 SHHCAGSLTNRWVYVTAHCFKSNMWDKPSLFSVLLGAWKLGSGPGRSKQVGIASVLP 131  
  
QY 108 Y---QDITADVALLKLSQVFTTSAILPICLPSTVKQLAIPPCFWCTGKVKKSSDRDY 164  
DB 132 YSWKEGTHADIALVRLERHSIQFSERILPICLPDSSVRLPPKTDICWAGWSIQDGVPLPH 191
```





---

This Page Blank (uspto)

Protease activity

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2001, 01:55:21 ; Search time 47.31 Seconds  
(without alignments)  
557.100 Million cell updates/sec

Title: US-09-854-844-2

Perfect score: 1863

Sequence: 1 MGPAACAFTELLLLGISVCG.....GRELTGEPLLTGLGDFIYNLK 346

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR-68:\*\*

1: pir1:\*\*

2: pir2:\*\*

3: pir3:\*\*

4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	518.5	27.8	343	1 A57014	proctasin (EC 3.4.21.1)
2	480	25.8	276	2 A38654	mast cell tryptase
3	479.5	25.7	270	2 S56160	mast cell tryptase
4	468.5	25.1	275	2 A32410	tryptase (EC 3.4.21.1)
5	468	25.1	274	2 JC4171	tryptase (EC 3.4.21.1)
6	462	24.8	275	2 A35863	tryptase (EC 3.4.21.1)
7	462	24.8	275	2 C35863	tryptase (EC 3.4.21.1)
8	461	24.7	275	2 B35863	tryptase (EC 3.4.21.1)
9	460.5	24.7	274	2 A45754	tryptase (EC 3.4.21.1)
10	459	24.6	273	2 A47246	tryptase (EC 3.4.21.1)
11	458.5	24.6	638	1 KQMSPL	plasma kallikrein
12	447.5	24.0	638	1 KQHUP	plasma kallikrein
13	446.5	24.0	638	1 KQRPPL	plasma kallikrein
14	445.5	23.9	625	1 KFHU1	coagulation factor
15	434.5	23.3	269	2 B32410	mastocytoma protease
16	429.5	23.1	1524	2 T30337	polyprotein - Afri
17	428	23.0	367	2 JE0104	testicular serine
18	414.5	22.2	366	2 JE0105	testicular serine
19	414.5	22.2	435	2 A61545	plasmin (EC 3.4.21.1)
20	414	22.2	245	1 KYBOA	chymotrypsin (EC 3.4.21.1)
21	413.5	22.2	460	2 B61545	plasmin (EC 3.4.21.1)
22	409	22.0	436	2 JX0172	acrosin (EC 3.4.21.1)
23	405	21.7	237	2 S68702	tryptase (EC 3.4.21.1)
24	404	21.7	416	1 S33777	hepsin (EC 3.4.21.1)
25	400.5	21.5	263	2 A31299	chymotrypsin (EC 3.4.21.1)
26	399.5	21.4	786	1 A47547	serine proteinase
27	399	21.4	264	2 I38136	chymotrypsin-like
28	397.5	21.3	417	1 S00845	hepsin (EC 3.4.21.1)
29	397	21.3	263	2 A21195	chymotrypsin (EC 3.4.21.1)

30	397	21.3	812	1 PLBO	plasmin (EC 3.4.21.1)
31	396.5	21.3	1035	1 A43090	enteropeptidase (E
32	394.5	21.2	790	1 PLPG	plasmin (EC 3.4.21.1)
33	393.5	21.1	810	1 PLRG	plasmin (EC 3.4.21.1)
34	392	21.0	812	1 PLMS	pancreatic elastase
35	390.5	21.0	271	1 ELRT2	enteropeptidase (E
36	390.5	21.0	1034	1 A53663	serine proteinase
37	389.5	20.9	253	2 A53968	acrosin (EC 3.4.21.1)
38	389.5	20.9	418	2 A37344	chymotrypsin (EC 3.4.21.1)
39	388	20.8	245	1 KYBOB	plasmin (EC 3.4.21.1)
40	387.5	20.8	810	2 B30848	serine proteinase
41	384.5	20.6	250	2 S55493	acrosin (EC 3.4.21.1)
42	383.5	20.6	415	1 A34170	chymotrypsin (EC 3.4.21.1)
43	382.5	20.5	263	1 KYRTB	pancreatic elastase
44	382.5	20.5	269	2 B26823	acrosin (EC 3.4.21.1)
45	382	20.5	421	1 S11674	acrosin (EC 3.4.21.1)

ALIGNMENTS

RESULT 1

A57014

proctasin (EC 3.4.21.1) precursor - human

C:Species: Homo sapiens (man)

C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 18-Jun-1999

C:Accession: A57014; A54866

R:Yu, J.X.; Chao, L.; Chao, J.

J. Biol. Chem. 270, 13483-13489, 1995

A:Title: Molecular cloning, tissue-specific expression, and cellular localization of

F:Reference number: A57014; MUID:95286644

A:Accession: A57014

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-343 <RES>

A:Cross-references: GB:D41351; NID:g862304; PIDN:AAC41759.1; PID:g862305

A:Experimental source: prostate

A:Note: parts of this sequence were determined by protein sequencing

R:Yu, J.X.; Chao, L.; Chao, J.

J. Biol. Chem. 269, 18843-18849, 1994

A:Title: Proctasin is a novel human serine proteinase from seminal fluid. Purification

A:Reference number: A54866; MUID:94308140

A:Accession: A54866

A:Molecule type: protein

A:Residues: 45-64 <YUA>

A:Genetics:

A:Gene: GDB:PRSS8

A:Cross-references: GDB:676446; OMIM:600823

A:Map position: 16p11.2-16p11.2

C:Superfamily: proctasin; trypsin homology

C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein

F:1-32/Domain: signal sequence #status predicted <Sig>

F:33-44,45-343/Product: proctasin #status predicted <NAT>

F:33-44/Domain: proctasin light chain #status predicted <CHL>

F:45-343/Domain: proctasin heavy chain #status predicted <CHH>

F:45-281/Domain: trypsin homology <TRY>

F:323-341/Domain: transmembrane #status predicted <TMM1>

F:37-154,70-86,168-242,201-223,234-262/Disulfide bonds: #status predicted

F:85,134,238/Active site: His, Asp, Ser #status predicted

F:159/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 27.8%; Score 518.5; DB 1; Length 343;

Best Local Similarity 37.4%; Pred. No. 1.5e-38;

Matches 123; Conservative 47; Mismatches 114; Indels 45; Gaps 10;

Qy 1 MGPAAC-AFTLLLLGI-----SVCGQPVYSSRVGGGODAAAGRPQVSLHED 48

Db 7 LGPQLGAVAILLYGLLRSGTGAGAEAPCG-VAPQRTGGSSVAGQMPQVSIYIE 65

Qy 49 HNFYTGSLVSERLILTAACIQPTTFTTWSYVWLGDSRRKRYKYSK-IVIHPR 107

Db 66 GVHVCGLSYSEQWVLSNAHCFPSHHKEAEVKGALHQLDYSYEDAKVSTLKDIIHPHS 125



A:Molecule type: mRNA  
A:Residues: 1-275 <V>  
A:Cross-references: GB:M24664; NID:g163982; PIDN:AAA30854.1; PID:g163983; GB:J02862  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase; zymogen  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-30/Domain: activation peptide #status predicted <ACT>  
F:31-275/Product: trypsin #status predicted <MAT>  
F:31-267/Domain: trypsin homology <TRY>  
F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 25.1%; Score 468.5; DB 2; Length 275;  
Best Local Similarity 37.5%; Pred. No. 3.3e-34;  
Matches 103; Conservative 50; Mismatches 95; Indels 27; Gaps 9;

QY 11 LLLLGISV-----CGQPVYSSRVGGQDAAGRWPMQVSLHFDHNF---IYGGSLVSERL 62  
DB 9 LALLGLVSPVSPAGQALQGVGVGREGAPGSKWPQVSLRLKQYWRHICGSLHPQW 68  
QY 63 ILTAACHCIPTWTFSTYVWLGSIYVGDGRKRVKY-----VSKIVHPKY--QDTTAD 114  
DB 69 VLTAACHVGN-----VVCPEIRV-QLREQHLYYQDHLPLPNRIVMPNYYTPENGAD 121  
QY 115 VALLKLSQVTFSTAILPICLPKSVTKQALPFPFCWVTGKGVKESDRDYHSALQAEYV 174  
DB 122 IALLELEDPPVNSAHVQVTLPPALQTFPTGTPCWTGWDVHSGTPLPPPLPKQVKVP 181  
QY 175 IDROAQEQLYPIGFIPLAEPVKEKICAGDTQNMKDSCKGSGGLPSCHIDQWITO 234  
DB 182 IVENSNCVQYH-LGLSTGDGVRIREDMLCAGNSKS--DSCQDGSGLVCRVRGWLQ 238  
QY 235 TGVVSGLECGK-SLPGVYTNVYIYQKWINATISR 268  
DB 239 AGVWSGEGCAQPNRGITRVAYILDHIHQIVPK 273

RESULT 5  
JC4171  
trypsin (EC 3.4.21.59) precursor - rat  
N:Alternate names: mast cell trypsinase  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 27-Aug-1995 #sequence\_revision 27-Oct-1995 #text\_change 20-Jun-2000  
C:Accession: JC4171  
R:Ida, H.; Itoh, H.; Tomita, M.; Murakumo, Y.; Kobayashi, T.; Maruyama, H.; Osada, Y.; N  
J. Biochem. 118, 210-215, 1995  
A:Title: cDNA sequencing and expression of rat mast cell trypsinase.  
A:Reference number: JC4171; MUID:96015171  
A:Accession: JC4171  
A:Molecule type: mRNA  
A:Residues: 1-274 <IDE>  
A:Cross-references: DBJ:D38455; NID:g556555; PIDN:BAA07486.1; PID:g556556  
C:Comment: This enzyme is basically specific for a connective tissue mast cell, it is up  
einate inhibitors.  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: glycoprotein; hydrolase; mast cell; serine proteinase; zymogen  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-29/Domain: activation peptide #status predicted <ACT>  
F:30-274/Product: mast cell trypsin #status predicted <MAT>  
F:30-266/Domain: trypsin homology <TRY>  
F:73,120,223/Active site: His, Asp, Ser #status predicted  
F:131/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 25.1%; Score 468; DB 2; Length 274;  
Best Local Similarity 36.6%; Pred. No. 3.6e-34;  
Matches 102; Conservative 53; Mismatches 84; Indels 40; Gaps 10;

QY 10 LLLLGISVCGQPVYS-----RVVGQDAAGRWPMQVSLHFDHNF---IYGGSLV 58  
DB 4 LLLLLALSPLASLVHAAPCPVQKRVGIVGREGASEKWPQVSLRPFKFSWHFHCGSLI 63  
QY 59 SERLIILTAACHCI-----QPTWTFSTYVWLGSIYVGDGRKRVKY-----VSKIVHPK 107

DB 64 HFQWLVTAACHVGLHIKSP-----LFRVQL-----REQVLYADQLLTNRNTRVVVHPH 111  
QY 108 YQ--DTTADVALLKLSQVTFSTAILPICLPKSVTKQALPFPFCWVTGKGVKESDRDYH 165  
DB 112 YTVEDGADIALLELEIPVNVSTHIPIPLPASETFPGSTSCWVTGWDGIDSDPEPLPP 171  
QY 166 SALQAEVPIIDROAQEQLYPIGFIPLAEPVKEKICAGDTQNMKDSCKGSGGLPS 225  
DB 172 YPLKQVKVPDIVENSCLDRKYH-TGLYTGDDVPVQDGMCLCAGNTRS--DSCQDGSGLV 228  
QY 226 CHIDGWIQTGVVSGLECGK-SLPGVYTNVYIYQKWIN 263  
DB 229 CRVKGTFWLAGVVSWGEGCAEARNPGITRVYVYLDWIH 267

RESULT 6  
A35863  
trypsin (EC 3.4.21.59) I precursor - human  
C:Species: Homo sapiens (man)  
C>Date: 12-Oct-1990 #sequence\_revision 12-Oct-1990 #text\_change 21-Jul-2000  
C:Accession: A35863; D35863; A60939; A39326  
R:Vanderslice, P.; Ballinger, S.M.; Ram, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey,  
Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990  
A:Title: Human mast cell trypsinase: multiple cDNAs and genes reveal a multigene serine  
A:Reference number: A35863; MUID:90251647  
A:Accession: A35863  
A:Molecule type: DNA  
A:Residues: 1-275 <V>  
A:Cross-references: GB:M33494; NID:g3927804; PIDN:AAC83172.1; PID:g339977  
A:Accession: D35863  
A:Molecule type: mRNA  
A:Residues: 1-275 <VA>  
A:Cross-references: GB:M33491  
R:Butterfield, J.H.; Weiler, D.A.; Hunt, L.W.; Wynn, S.R.; Roche, P.C.  
J. Leukoc. Biol. 47, 409-419, 1990  
A:Title: Purification of trypsinase from a human mast cell line.  
A:Reference number: A60939; MUID:90244210  
A:Accession: A60939  
A:Molecule type: protein  
A:Residues: 31-38, 'P', 40-41, 'X', 43, 'T', 45-48, 'X', 50 <BUT>  
A:Experimental source: mast cell  
A:Note: 44-Gly was also found  
R:Cromlish, J.A.; Seidah, N.G.; Marcinkiewicz, M.; Hamelin, J.; Johnson, D.A.; Chretel  
J. Biol. Chem. 262, 1363-1373, 1987  
A:Title: Human pituitary trypsinase: molecular forms, NH-2-terminal sequence, immunocyt  
A:Reference number: A39326; MUID:87109258  
A:Accession: A39326  
A:Molecule type: protein  
A:Residues: 31-38 <CRO>  
A:Experimental source: pituitary  
C:Genetics:  
A:Introns: 21/1; 78/2; 177/1; 221/3  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase; zymogen  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-30/Domain: activation peptide #status predicted <ACT>  
F:31-275/Product: trypsin #status experimental <MAT>  
F:31-267/Domain: trypsin homology <TRY>  
F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 24.8%; Score 462; DB 2; Length 275;  
Best Local Similarity 36.9%; Pred. No. 1.2e-33;  
Matches 104; Conservative 50; Mismatches 82; Indels 46; Gaps 12;

QY 11 LLLLGISV-----GQPVYSSRVGGQDAAGRWPMQVSL-----HFDHNF 52  
DB 4 LLLLLALPVLASRAYAAPAFQALQVRGIVGQEPAPSKWPQVSLRVHGVPMWHF----- 58  
QY 53 YGSLVSERLIILTAACHCIPTWTFSTYVWLGSIYVGDGRKRVKY-----VSKIVHP 106  
DB 59 CGSLIHPQWLVTAACHVGPDKVD-----LAALRV-QLREQHLYYQDQLLPVSRIVHP 111





QY 112 TADVALLKLSOVTTTSAILPLCLPSVTQLAIPPCFQWVGWKVKESDRDYSALOEAE 171  
Db 481 NYDIALIKLOTLPLNTEFOKPICLPSKADNTIYNCWWTGMYTKEGET--QNILQA 538  
QY 172 EVDPIRDAEQEOLNPIGIFLPALEPVFKEDKICAGDTQNMKDSCGSGGPLSCHIDGV 231  
Db 539 TIPLVPNEECQKKR-----DYVINQMICAGYKEGGTDACKGDSGGPLCKVHSGR 589

QY 232 WIQTGVWSGLECG-KSLPGVYTNVIYYOKWI 262  
Db 590 WLVGITSWGEGCRKRDPGVYTKVSEYMDWI 621

RESULT 12  
KOUP  
plasma kallikrein (EC 3.4.21.34) precursor - human  
N:Alternate names: kininogenin; plasma prekallikrein  
C:Species: Homo sapiens (man)  
C>Date: 13-Aug-1986 #sequence\_revision 13-Aug-1986 #text\_change 18-Jun-1999  
C:Accession: A00921; A37939  
R:Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.  
Biochemistry 25, 2410-2417, 1986  
A:Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four b  
A:Reference number: A00921; MUID:86243359  
A:Accession: A00921  
A:Molecule type: mRNA  
A:Residues: 1-638 <CHU>  
A:Cross-references: GB:M13143; NID:g190262; PIDN:AAA60153.1; PID:g190263  
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.  
Biochemistry 30, 2050-2056, 1991  
A:Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of  
A:Reference number: A37939; MUID:91152016  
A:Accession: A37939  
A:Molecule type: protein  
A:Residues: 20-27;40-46,'X','48','H';50,'X';52-70,'H';75-76,'X';78-80;103-113;131-140;141-  
260-283,'X';285;287-291,'X';293-295;314-317,'X';319-320;321-324,'X';329-333;334-339,'X'  
525;538-551;562,'X';564-567;573,'X';575-576;578-583,'X';585;592-604 <MCN>  
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w  
C:Comment: The zymogen is activated by factor XIa, which cleaves the molecule into a li  
are linked by one or more disulfide bonds.  
C:Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal r  
inogen and may also play a role in the renin-angiotensin system by converting prorenin i  
C:Genetics:  
A:Gene: GDB:KLK3  
A:Cross-references: GDB:l27575; OMIM:229000  
A:Map position: 4q35-q435  
C:Superfamily: coagulation factor XI; trypsin homology  
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla  
F:1-19/DNA: signal sequence #status predicted <SIG>  
F:20-638/Product: plasma kallikrein #status predicted <SIG>  
F:20-390/DNA: plasma kallikrein heavy chain #status predicted <HCH>  
F:20-109/DNA: apple repeat <AP1>  
F:110-199/DNA: apple repeat <AP2>  
F:200-289/DNA: apple repeat <AP3>  
F:291-380/DNA: apple repeat <AP4>  
F:391-638/DNA: plasma kallikrein light chain #status predicted <LCH>  
F:391-638/DNA: tryptsin homology <TRY>  
F:21-104,47-51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,322-328,383  
F:127,308,396,453,494/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:318-347,340-345/Disulfide bonds: #status predicted  
F:390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted  
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 24.0%; Score 447.5; DB 1; Length 638;  
Best Local Similarity 36.4%; Pred. No. 7.1e-32;  
Matches 99; Conservative 43; Mismatches 103; Indels 27; Gaps 8;

QY 2 GPACCAFTLLLLLGISVCGOPVSVSRVGGODAAAGRWPMQVSLHFD---HNFIYGGSVLV 58  
Db 366 GSSGYSLRLCTNGDNSVCTTKT-STRIVGGTNSSGEPMQVSLVQLKTAQRHLCCGSLI 424

QY 59 SERLILTAHC-----IQPTWTFYTVWLGSITVCDGRKRVKYVSKIVHPKYO--DT 111  
Db 425 GHOWYLTAACHFDGLPQDWIRYSGILNLSDITKDTFPQIK----EIHHQNYKVSEG 480  
QY 112 TADVALLKLSOVTTTSAILPLCLPSVTQLAIPPCFQWVGWKVKESDRDYSALOEAE 171  
Db 481 NYDIALIKLOTLPLNTEFOKPICLPSKADNTIYNCWWTGMYTKEGET--QNILQA 538  
QY 172 EVDPIRDAEQEOLNPIGIFLPALEPVFKEDKICAGDTQNMKDSCGSGGPLSCHIDGV 231  
Db 539 TIPLVPNEECQKKR-----DYVINQMICAGYKEGGTDACKGDSGGPLCKVHSGR 589

QY 232 WIQTGVWSGLECG-KSLPGVYTNVIYYOKWI 262  
Db 590 WLVGITSWGEGCARREQBPQGVYTKVAEYMDWI 621

RESULT 13  
KORTPL  
plasma kallikrein (EC 3.4.21.34) precursor - rat  
N:Alternate names: Fletcher factor; kininogenin; serum kallikrein  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 18-Jun-1999  
C:Accession: A39180; A33320; S06851; I53041; S06852  
R:Beaubien, G.; Rosinski-Chupin, I.; Mattel, M.G.; Mbikay, M.; Chretien, M.; Seidah,  
Biochemistry 30, 1628-1635, 1991  
A:Title: Gene structure and chromosomal localization of plasma kallikrein.  
A:Reference number: A39180; MUID:91129236  
A:Accession: A39180  
A:Molecule type: DNA  
A:Residues: 1-638 <BEA>  
A:Cross-references: GB:J05315  
A>Note: the authors translated the codon GAG for residue 81 as Glu  
R:Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Roug  
DNA 8, 563-574, 1989  
A:Title: The cDNA structure of rat plasma kallikrein.  
A:Reference number: A33320; MUID:90091743  
A:Accession: A33320  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-638 <SER>  
A:Cross-references: GB:M30282; NID:g205010; PIDN:AAA1463.1; PID:g205011  
A>Note: part of this sequence, including the amino ends of both the heavy  
R:Napkin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah,  
Biochim. Biophys. Acta 999, 103-110, 1989  
A:Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and  
A:Reference number: S06851; MUID:90089457  
A:Accession: S06851  
A:Molecule type: protein  
A:Residues: 20-45;391-413 <PAO>  
R:Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Roug  
DNA Cell Biol. 8, 563-574, 1989  
A:Title: The cDNA structure of rat plasma kallikrein.  
A:Reference number: I53041  
A:Accession: I53041  
A>Status: translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-638 <RES>  
A:Cross-references: GB:M58590; NID:g206721; PIDN:AAA2069.1; PID:g206722  
C:Comment: This protein, synthesized in the liver, circulates as a noncovalen  
C:Comment: The zymogen is activated by factor XIa, which cleaves the molecula  
are linked by one or more disulfide bonds.  
C:Genetics:  
A:Gene: PK  
C:Superfamily: coagulation factor XI; trypsin homology  
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hy  
F:1-19/DNA: signal sequence #status predicted <SIG>  
F:20-390/Product: plasma kallikrein heavy chain #status experimental <MA  
F:20-109/DNA: apple repeat <AP1>  
F:110-199/DNA: apple repeat <AP2>  
F:200-289/DNA: apple repeat <AP3>  
F:291-380/DNA: apple repeat <AP4>  
F:391-638/Product: plasma kallikrein light chain #status experimental <MA



A;Residues: 1-269 <VAN>  
A;Cross-references: GB:M24665; NID:gl63984; PID:gl63985; GB:J02862  
A;Note: the authors translated the codon ACG for residue 114 as Ser  
C;Superfamily: trypsin; trypsin homology  
C;Keywords: hydrolase; serine proteinase  
F;1-19/Domain: signal sequence #status predicted <SIG>  
F;20-269/Product: mastocytoma proteinase #status predicted <MAT>  
F;20-259/Domain: trypsin homology <TRY>  
F;66,116,217/Active site: His, Asp, Ser #status predicted

Query Match 23.3%; Score 434.5; DB 2; Length 269;  
Best Local Similarity 40.4%; Pred. No. 3.4e-31;  
Matches 103; Conservative 36; Mismatches 89; Indels 27; Gaps 10;

```
QY 28 VYGGDAAAGRWPQVSLHF-----DHNFIYGGSLVSRILITAAHCIQ-PTWTTFST 80
Db 20 IVGGCKVPARRYPWQVSLRFHGMGSGQWQHICGSLIHPQWVLTAAHCVLEGLEAATLR 79
QY 81 VMLGSTVGDSTRKRVKYYVSKIYHPKYODT-----TADVALLKSSQVTFTSAILPICL 135
Db 80 VQVQLRLYDHDQLCN--VTEIRHPNFNMSWYGMTADIALLKLEAPLTLSDEVNLVSL 137
QY 136 PSVTKQLATPP--FCWVTGKVKKESS--DRDYHSALQEAEPITDROACEOLYNPIGIF 191
Db 138 PS--PSLIYPPGMLCWVTGWDIADHTPLPPYH--LQEVETIVGNRECNCHYQTI--- 190
QY 192 LPALPEVIKEDKICAGDTQNMKDSCKGDSGGPLSCHIDGWIQTGVVSWGLECGKSLPGV 251
Db 191 LEQDDEVIIKQDMLCAGSEGH--DSCQMDSGGPLVCHWKCTWIQGVVSWGGYGNLPGV 248
QY 252 YTNVIYYQKWINATI 266
Db 249 YARVTSYVSWIHQHI 263
```

Search completed: November 22, 2001, 02:39:28  
Job time: 2647 sec



```
FT PROPEP 30 32 ACTIVATION PEPTIDE.
FT CHAIN 33 44 PROSTATIN LIGHT CHAIN.
FT CHAIN 45 322 PROSTATIN HEAVY CHAIN.
FT PROPEP 323 343
FT TRANSEM 320 340 POTENTIAL.
FT DISULFID 37 154 INTERCHAIN (BY SIMILARITY).
FT DISULFID 70 86 BY SIMILARITY.
FT DISULFID 168 244 BY SIMILARITY.
FT DISULFID 201 223 BY SIMILARITY.
FT DISULFID 234 262 BY SIMILARITY.
FT CARBOHYD 159 159 N-LINKED (GLCNAC... ) (POTENTIAL).
FT ACT_SITE 85 85 CHARGE RELAY SYSTEM.
FT ACT_SITE 134 134 CHARGE RELAY SYSTEM.
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM.
FT SEQUENCE 343 AA; 36431 MW; 98DD6447F5A8C1B2 CRC64;

Query Match 27.8%; Score 518.5; DB 1; Length 343;
Best Local Similarity 37.4%; Pred. No. 9.2e-38;
Matches 123; Conservative 47; Mismatches 114; Indels 45; Gaps 10;

QY 1 MGPAGC-AFTLLLLGI-----SYCGQPVYSSRVVGGQDRAAGRWQVSLHFD 48
DB 7 LGPEQLGAVAILLYGLRSGTGAEGAEPCG-VAPOARITGGSSAVAGQMPQVSIYE 65

QY 49 HNFYGGSLVSEILITAAHCIOPTTTFSTVWLGSITVDSKRKVKYVSK-IVIHPK 107
DB 66 GVHVCGGSLVSEQWLSNAHCFFSEHKEAYEVLKGNHQDLSYSDAKVSTLKDIPHPS 125

QY 108 Y--QDTTADVALKLKLSQVFTTSAILPCLPSVTQKLAIPFPFCWVTGWKVKSSDRDYH 165
DB 126 YLQBSGSDIALLQLSPITFSRVIRPCLPAANASPPNGLHCTVTGNGHVPASVSLTLP 185

QY 166 SALQAEVPIIDRQACQLNPIGIFLPALEP-VIKEDKICAGDTQNKDCKGDSGGPL 224
DB 186 KPLQGLEVPLISRETCLN---IDAKPEEPHFVQEDMYCAGYVEGGKDACQDGGSL 242

QY 225 SCHIDGVIQTVGVVSWGLECG-KSLPGVNTVYQKINATISR----- 268
DB 243 SCPVEGLWYLTGVSGWDGACGARNRPGVYTLASSYASWIOSKVTELOPRVVPQTOESQPD 302

QY 269 ----ANLDFSD-----FLPPIVLLSLAL 288
DB 303 SNLCGSHLAFSSAPAQGLLRPLFLPLGL 331

RESULT 2
MCT6_MOUSE STANDARD; PRT; 276 AA.
AC P21845; Q61962;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MAST CELL PROTEASE 6 PRECURSOR (EC 3.4.21.-) (MMP-6) (TRYPTASE).
GN MCT6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=91139682; PubMed=1995638;
RA Reynolds D.S., Gurley D.S., Austen K.F., Serafin W.E.;
RT "Cloning of the cDNA and gene of mouse mast cell protease-6.
RT Transcription by progenitor mast cells and mast cells of the
RT connective tissue subclass.";
RL J. Biol. Chem. 266:3847-3853(1991).
[2]
RN SEQUENCE FROM N.A.
RX STRAIN=LEADEN X A1;
RX MEDLINE=94023807; PubMed=8210998;
RA Huang R., Abbrink M., Gobl A.E., Nilsson G., Aveskogh M., Larsson L.G.,
RA Nilsson K., Hellman L.;
```

```
RT "Expression of a mast cell tryptase in the human monocytic cell lines
RT U-937 and Mono Mac 6.";
RT Scand. J. Immunol. 38:359-367(1993).
[3]
RN SEQUENCE FROM N.A. (SHORT FORM).
RP STRAIN=LEADEN X A1;
RC MEDLINE=95048382; PubMed=7959952;
RX Huang R., Hellman L.T.;
RA "Genes for mast-cell serine protease and their molecular evolution.";
RL Immunogenetics 40:397-414(1994).
[4]
RN SEQUENCE OF 32-54.
RP MEDLINE=90222202; PubMed=2326280;
RX Reynolds D.S., Stevens R.L., Lane W.S., Carr M.H., Austen K.F.,
RA Serafin W.E.;
RT "Different mouse mast cell populations express various combinations
RT of at least six distinct mast cell serine proteases.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:3230-3234(1990).
CC -I- ALTERNATIVE PRODUCTS: TWO FORMS OF MCT6 ARE PRODUCED BY
CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SHORT FORM IS PROBABLY
CC A NON FUNCTIONAL VARIANT.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. STRONGEST TO OTHER GRANZYMES AND TO MAST CELL
CC PROTEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M57626; AAA39988.1; -
CC EMBL; M57625; AAA39987.1; -
CC EMBL; L31853; AAA39725.1; -
CC EMBL; X78542; CAA55288.1; -
CC PIR; A38654; D38654.
CC HSSP; P20231; 1AAO.
CC MEROPS; S01.025; -
CC MGD; MGI:96942; Mct6.
CC InterPro; IPR001254; -
CC InterPro; IPR001314; -
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;
KW Alternative splicing.
FT SIGNAL 1 21 ACTIVATION PEPTIDE.
FT PROPEP 22 31 MAST CELL PROTEASE 6.
FT CHAIN 32 276 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 122 122 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 225 225 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 60 76 BY SIMILARITY.
FT DISULFID 156 231 BY SIMILARITY.
FT DISULFID 189 212 BY SIMILARITY.
FT DISULFID 221 249 BY SIMILARITY.
FT CARBOHYD 133 133 N-LINKED (GLCNAC... ) (POTENTIAL).
FT VARSPLIC 223 230 GDSGGPLV -> PFCIGDDI (IN SHORT ISOFORM).
FT VARSPLIC 231 276 MISSING (IN SHORT ISOFORM).
FT SEQUENCE 276 AA; 30927 MW; 525B2C9A04A72200 CRC64;
SQ

Query Match 25.8%; Score 480; DB 1; Length 276;
Best Local Similarity 37.8%; Pred. No. 1.6e-34;
Matches 104; Conservative 51; Mismatches 88; Indels 32; Gaps 9;

QY 10 LLLLLGIVCGQPVYSS-----RVVGGQDAAAGRWQVSLHFDHNF---YGGSLV 58
DB 6 LLLWALSLASLVISAPRANQRVGVGGHEASESKWPQVSLRFLKNYWHFCGSLI 65
```

```

FT CARBOHYD      228      228      N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE     270 AA; 30166 MW; 1BE102DB86943401 CRC64;

Query Match          25.7%; Score 479.5; DB 1; Length 270;
Best Local Similarity 35.2%; Pred No. 1.7e-34;
Matches 101; Conservative 51; Mismatches 84; Indels 51; Gaps 10;

QY 11 LLLLGISVCGQPVSRR-----VVGSDAAAGRWPMQVSLHFDHNF---IYGSSL 57
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 4 LLLLAL-----PLFSLMRSPLCQEWGILGVSGEAPGNKWPQVSLRANETWRRHFCGSSL 58

QY 58 VSERLITAHACTOPTWTTSYTWLGSITVGDS-----RKRVKY-----VSKIIV 104
   | : ||||| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 59 IHPQWLTAACHVGP-----TIADPNKVRVOLRKOYLXYHDHLAVSRIIT 104

QY 105 HPKYQDTT--ADVALLKSQVTFTSAILLPTCLPSVTKQLAIPPCQVWGKGKVESDDR 162
   || : || : ||||| : | : | : | : | : | : | : | : | : | : | :
Db 105 HPIFYATONGADIALLEKPNISSHVHVPSPASETPPGTLCTWGTGMGINDNVSL 164

QY 163 DYHSALQAEVIPIDROAQCELYNIGIFLPALEPVKEIKDACGDTONMKDSCKGSGG 222
   -|-|-|-|-|-|-| : | : | : | : | : | : | : | : | : | : | :
Db 165 PPPPLEKVQPVVENQLCDLKYYHK-GVYTGDNTHIVRDMLCAGNEG--DSQGDSGG 221

QY 223 PLSCHIDGVMIQTGVSVWGLEGCG-KSLPGVYTNVIYOKWINATISR 268
   || : | : ||||| : | : | : | : | : | : | : | : | : | : | :
Db 222 PLVCKVNGTWLQAGVSWGEGCALPNRPGIYTRVITYLVLDWIHRVVKP 268

RESULT 4
TRYT_CANFA STANDARD; PRG; 275 AA.
ID TRYT_CANFA ID TRYT_CANFA
AC PI5944;
DC 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE TRYTOUSE PRECURSOR (EC 3.4.21.59).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
ON NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89352460; PubMed=2504277;
RT Vanderslice P., Craik C.S., Nadel J.A., Caughey G.H.;
RA "Molecular cloning of dog mast cell tryptase and a related protease;
RT structural evidence of a unique mode of serine protease activation.";
RL Biochemistry 28:4148-4155(1989).
CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST
CC CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION
CC RESPONSE OF THIS CELL TYPE.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-|- , LYS-|- , BUT
CC WITH MORE RESTRICTED SPECIFICITY THAN TRYPsin.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON
CC MAST CELL ACTIVATION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. STRONGEST TO OTHER TRYPTASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24664; AAA30854.1; -.
CC PIR; A32410; A32410.
CC HSP; P20231; IAAO.
CC MEROPS; S01.143; -.
CC InterPro; IPR001254; -.
CC InterPro; IPR001314; -.

```



DE ALPHA-TRYPTASE PRECURSOR (EC 3.4.21.59) (TRYPTASE 1).  
GN TP51.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=90009311; PubMed=2677049;  
RA Miller J.S., Westin E.H., Schwartz L.B.;  
RT "Cloning and characterization of complementary DNA for human  
RT tryptase.";  
RL J. Clin. Invest. 84:1188-1195(1989).  
RN [2]  
RP REVISIONS TO 89-93 AND 108.  
RC TISSUE=Lung;  
RA Schwartz L.B.;  
RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 31-50, AND PITUITARY, SEQUENCE OF 31-38.  
RC TISSUE=Lung;  
RX MEDLINE=87109258; PubMed=3543004;  
RA Cromlish J.A., Seidah N.G., Marcinkiewicz M., Hamelin J., Johnson D.A.,  
RA Chretien M.;  
RT "Human pituitary tryptase: molecular forms, NH2-terminal sequence, and  
RT immunocytochemical localization, and specificity with prohormone and  
RT fluorogenic substrates.";  
RL J. Biol. Chem. 262:1363-1373(1987).  
CC -!- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
CC CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION  
CC RESPONSE OF THIS CELL TYPE.  
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-|-, LYS-|-, BUT  
CC WITH MORE RESTRICTED SPECIFICITY THAN TRYPSIN.  
CC -!- SUBUNIT: HOMOTETRAMER.  
CC -!- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON  
CC MAST CELL ACTIVATION.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY. STRONGEST TO OTHER TRYPTASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M30038; AA86934.1; -.  
DR PIR; A45754; A45754.  
DR HSP; P20231; 1AAO.  
DR MEROPS; S01.143; -.  
DR MIM; 191080;  
DR InterPro; IPR001254; -.  
DR InterPro; IPR001314; -.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease; Signal; Glycoprotein.  
FT SIGNAL 1 30  
FT CHAIN 31 275  
FT ACT\_SITE 74 74  
FT ACT\_SITE 121 121  
FT ACT\_SITE 224 224  
FT DISULFID 59 75  
FT DISULFID 155 230  
FT DISULFID 188 211  
FT DISULFID 220 248  
FT CARBOHYD 132 132  
FT CARBOHYD 233 233  
FT SEQUENCE 275 AA; 30730 MW; BDAAC4B8CF96CD71 CRC64;

Query Match 25.1%; Score 467; DB 1; Length 275;  
Best Local Similarity 37.2%; Pred. No. 2.le-33;  
Matches 103; Conservative 50; Mismatches 88; Indels 36; Gaps 11;  
QY 11 LLLLSVCGQPVS-----SRVVGQDAARWQVSLHFDHNF---YGGSL 57  
DB 4 LLLLALPVLASRAYAPVQALQQAGIVGGGAPSKWQVSLRVDRYWNHFCGSL 63  
QY 58 VSERLILTAHCIOPTWTFSTVWLGSITVGDSSRRKRYI-----VSKIVHPKYQ-- 109  
DB 64 IIPQWVLTAAHCLGPDVKD-----LATLRV-QLRQHLHYQDQLLPVSRIVHPQFYII 116  
QY 110 DTTADVALLKSSQVTFTSAILPLICLPVTKQLAIPPF--CWWTGKVKESDRDYHSA 167  
DB 117 QTGADIALLELEPEVNISSRVHTVMLPPASE--TFPPGPGCWVTGWDVNDLEPLPPFP 174  
QY 168 LQAEVPIIDRQACEQLYNPIGIFLPALEPVKEDKICAGDTQNMKDKSGDGGPLSCH 227  
DB 175 LKQVKVPIMHENHICDAKYH-LGAYTGDDVRIIRDDMLCAGNSQ--RDSCKGDSGGLVCK 231  
QY 228 IDGVNIQTGVSWGLECGK-SLPGVVTNVIYKQWIN 263  
DB 232 VNGTWLQAGVSVNDEGCAQPNRPRIYTRYIYLDWIH 268  
RESULT 7  
TRYB\_HUMAN  
ID TRYB\_HUMAN STANDARD; PRT; 275 AA.  
AC P20231;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE BETA-TRYPTASE PRECURSOR (EC 3.4.21.59) (TRYPTASE 2).  
GN TP5B1 OR TPS2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=90369005; PubMed=2203827;  
RA Miller J.S., Moxley G., Schwartz L.B.;  
RT "Cloning and characterization of a second complementary DNA for human  
RT tryptase.";  
RL J. Clin. Invest. 86:864-870(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93166209; PubMed=8434231;  
RA Blom T., Hellman L.;  
RT "Characterization of a tryptase mRNA expressed in the human basophil  
RT cell line K0812.";  
RL Scand. J. Immunol. 37:203-208(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90251647; PubMed=2187193;  
RA Vanderslice P., Ballinger S.M., Tam E.K., Goldstein S.M., Craik C.S.,  
RA Caughey G.H.;  
RT "Human mast cell tryptase: multiple cDNAs and genes reveal a  
RT multigenic serine protease family.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:3811-3815(1990).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).  
RX MEDLINE=98180625; PubMed=9521329;  
RA Pereira P.J.B., Bergner A., Macedo-Ribeiro S., Huber R.,  
RA Matschner G., Fritz H., Sommerhoff C.P., Bode W.;  
RT "Human beta-tryptase is a ring-like tetramer with active sites facing  
RT a central pore.";  
RL Nature 392:306-311(1998).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (3.4 ANGSTROMS).  
RX MEDLINE=99432168; PubMed=10500112;





FT CHAIN 29 273 MAST CELL PROTEASE 7.  
FT ACT\_SITE 72 72 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 119 119 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 222 222 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 57 73 BY SIMILARITY.  
FT DISULFID 153 228 BY SIMILARITY.  
FT DISULFID 186 209 BY SIMILARITY.  
FT DISULFID 218 246 BY SIMILARITY.  
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPLIC 44 46 VSL -> GCC (IN TRUNCATED ISOFORM C57BL/6J).  
FT VARSPLIC 47 273 MISSING (IN TRUNCATED ISOFORM C57BL/6J).  
SQ SEQUENCE 273 AA; 30337 MW; 50ECB4765294205E CRC64;

Query Match 24.6%; Score 459; DB 1; Length 273;  
Best Local Similarity 35.9%; Pred. No. 1e-32;  
Matches 101; Conservative 49; Mismatches 85; Indels 46; Gaps 10;

QY 11 LLLLGISVCGQPYSS-----RVVGGDAAAGRWPMQVSLHDFHNF---YVGSLSVS 59  
DQ 4 LLLLTPLLSLVHAAPGPAWTRGVGGQEAHGNKWPQVSLRANDTYWVHFCGSLIH 63  
QY 60 ERLILAAHCIOPTWTFSTVNLGTSIVGDS-----RKRKYK-----VSKVIVHP 106  
DQ 64 POWLTAACHVGP-----DVAADPNKRVQLRQVLYHDHLMVTSQIITHP 109  
QY 107 KY---ODTADVALLKLSOVTFSTAILPLPSVTKQLAIPFPFCWVTKGKVKESDRD 163  
DQ 110 DFVIVQD-GADIALKLKTNPNVSDVHPVLPSPAKETFPSTGLCWVTKGMDNGVNL 168  
QY 164 YHSAQAEYPIIDROACEOLYPIGIFLPALEPVKEDKICAGDTQNMKDSCKGSGGP 223  
DQ 169 PPPLKEVQVPIIENHLCDLKLYHK-GLITGDNVHIVRDDMLCAGNEGH---DSQGDGSGP 225  
QY 224 LSCHDGWIQTGVSWGLECGK-SLPGVTNVIYKQWIN 263  
DQ 226 LVCKVEDTQLQAGVSWGEGCAQPNRPFGIYTRYVYLDWTH 266

RESULT 9  
KAL\_MOUSE STANDARD; PRT; 638 AA.  
AC P26262;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)  
DE (KININOGENIN) (FLETCHER FACTOR).  
GN K13 OR PK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN-BALB/C; TISSUE=Liver;  
RX MEDLINE=91090844; PubMed=2264928;  
RA Seidman N.G., Sawyer N., Hamelin J., Mion P., Beaubien G.,  
RA Braichpapa L., Rochement J., Mbikay M., Chretien M.;  
RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,  
RT and comparison of protein and mRNA levels among species.";  
RL DNA Cell Biol. 9:737-748(1990).  
CC -!- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT  
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING  
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM  
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN  
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.  
CC -!- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES  
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,  
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE  
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.

CC -!- SIMILARITY: CONTAINS 4 APPLE DOMAINS.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; MS8588; AAA63393.1; -.  
DR PIR; A36557; KOMSPL.  
DR HSP; P00750; LRTE.  
DR MEROPS; S01.212; -.  
DR MGD; MGI:102849; K1k3.  
DR InterPro; IPR000177; -.  
DR InterPro; IPR001254; -.  
DR InterPro; IPR001314; -.  
DR InterPro; IPR003014; -.  
DR Pfam; PF00024; PAN; 4.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00005; APPLIEDOMAIN.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
DR PROSITE; PS00495; APPLE; 4.  
DR Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;  
KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;  
KW Repeat.  
FT SIGNAL 1 19  
FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.  
FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.  
FT DOMAIN 20 105 APPLE 1.  
FT DOMAIN 110 195 APPLE 2.  
FT DOMAIN 200 285 APPLE 3.  
FT DOMAIN 291 376 APPLE 4.  
FT DOMAIN 389 621 CATALYTIC.  
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 453 453 PROBABLE.  
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT ACT\_SITE 434 434 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 483 483 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 578 578 CHARGE RELAY SYSTEM.  
FT DISULFID 21 104 BY SIMILARITY.  
FT DISULFID 47 77 BY SIMILARITY.  
FT DISULFID 51 57 BY SIMILARITY.  
FT DISULFID 111 194 BY SIMILARITY.  
FT DISULFID 137 166 BY SIMILARITY.  
FT DISULFID 141 147 BY SIMILARITY.  
FT DISULFID 201 284 BY SIMILARITY.  
FT DISULFID 227 256 BY SIMILARITY.  
FT DISULFID 231 237 BY SIMILARITY.  
FT DISULFID 292 375 BY SIMILARITY.  
FT DISULFID 318 347 BY SIMILARITY.  
FT DISULFID 322 328 BY SIMILARITY.  
FT DISULFID 340 345 BY SIMILARITY.  
FT DISULFID 383 503 BY SIMILARITY.  
FT DISULFID 419 435 BY SIMILARITY.  
FT DISULFID 517 584 BY SIMILARITY.  
FT DISULFID 548 563 BY SIMILARITY.  
FT DISULFID 574 602 BY SIMILARITY.  
SQ SEQUENCE 638 AA; 71368 MW; CC27C93AB1086599 CRC64;

Query Match 24.6%; Score 458.5; DB 1; Length 638;  
Best Local Similarity 36.4%; Pred. No. 3e-32;  
Matches 99; Conservative 42; Mismatches 104; Indels 27; Gaps 8;  
QY 2 GPAGCAFTLLLLGISVCGQPYSSRVVGGDAAAGRWPMQVSLH---FDHNFYGGSLV 58



```
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U75329; AAC51784.1; -.
DR HSP; P00763; LDPO.
DR MIM; 602060; -.
DR InterPro; IPR001254; -.
DR InterPro; IPR001314; -.
DR InterPro; IPR002172; -.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; BR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00668; LDLRA_2; 1.
DR Hydrolase; Serine protease; Transmembrane; Signal-anchor.
FT DOMAIN 1 84 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 85 105 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 106 492 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 112 150 LDL-RECEPTOR CLASS A.
FT DOMAIN 151 243 SRCR.
FT DOMAIN 256 492 CATALYTIC.
FT ACT_SITE 296 296 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 441 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 113 126 BY SIMILARITY.
FT DISULFID 120 139 BY SIMILARITY.
FT DISULFID 133 148 BY SIMILARITY.
FT DISULFID 244 365 BY SIMILARITY.
FT DISULFID 281 297 BY SIMILARITY.
FT DISULFID 410 426 BY SIMILARITY.
FT DISULFID 437 465 BY SIMILARITY.
FT CARBOHYD 213 213 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 492 AA; 53847 MW; 7EAAFD18609DDA CRC64;
CC -----
Query Match 24.4%; Score 455.5; DB 1; Length 492;
Best Local Similarity 35.8%; Pred. No. 4e-32;
Matches 98; Conservative 39; Mismatches 100; Indels 37; Gaps 8;
QY 6 CAPTLLLLIGISVCGQPVVS---SRVVGQDAAAGRWQVSLFPHNFIYGGSLVSERL 62
DB 231 CSKAVYSLRCLACGYNLSRSRQISVGGESALPGAMPQVSLHVQVHVCVGSIIPTPEW 290
QY 63 ILTAACHIOPT-----WTFSTVVLGSIITVGDSSKRVKYY-----VSKIVTHRPYQDT 111
DB 291 IVTAACHVEKPLNPNHWTAF-----GILQSFYFAGYQVQKVISHPNWDK 340
QY 112 TA--DVALLKSSQVFTTSAILPCLPSVTQKLAIPFFCVWTVGWKVKSSDRDYHSALQ 169
DB 341 TKNNIDIALMKIQKPLFDNLVQVCLPNPQGMQLPEQLCWISGWGATEEKGKTS--EVLN 398
QY 170 EAEVPIIDRQACQLNYPNGIFLPALEPVKEKICAGDTQNNMKDSKGDGSGPLSCHID 229
DB 399 AAKVLLIETQRCNSRYVDNLITPAM-----ICAGFLQGNVDSQCGDSGGLVTSNN 450
QY 230 GVIQIQTGVVSWGLECKSL-PGYVTNVIYQKWI 262
DB 451 NIWMLIGDTSWGCGCAKAYRPGVYGVNMVFTDWI 484
```

```
RESULT 12
TMS2_MOUSE
ID TMS2_MOUSE STANDARD; PRT; 490 AA.
AC Q9JQ08; Q9JKC4; Q9Y82;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-) (EPITHELIALIN) (PLASMIC
DE TRANSMEMBRANE PROTEIN X).
GN TMPRSS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viikko P.T.;
RT "Expression of transmembrane serine protease TMPRSS2 in mouse and
RT human tissues.";
RL J. Pathol. 191:0-0(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Han J., Kim S.;
RT "Putative transmembrane protease X.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA Jacquinet E.J., Rao N.V., Rao G.N., Hoidal J.R.;
RT "A novel mosaic serine protease, epithelialin.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF199362; AAF97867.1; -.
DR EMBL; AF243500; AAF64186.1; -.
DR EMBL; AF113596; AAF21308.1; -.
DR HSP; P00761; IAKS.
DR MGD; MGI:1354381; Tmprss2.
DR InterPro; IPR001254; -.
DR InterPro; IPR001314; -.
DR InterPro; IPR002172; -.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00668; LDLRA_2; 1.
DR Hydrolase; Serine protease; Transmembrane; Signal-anchor.
FT DOMAIN 1 83 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 84 104 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 105 490 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 111 149 LDL-RECEPTOR CLASS A.
FT DOMAIN 150 242 SRCR.
FT DOMAIN 254 490 CATALYTIC.
FT ACT_SITE 294 294 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 343 343 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 439 439 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 112 125 BY SIMILARITY.
FT DISULFID 119 138 BY SIMILARITY.
```

FT	DISULFID	132	147		BY SIMILARITY.
FT	DISULFID	243	363		BY SIMILARITY.
FT	DISULFID	279	295		BY SIMILARITY.
FT	DISULFID	408	424		BY SIMILARITY.
FT	DISULFID	435	463		BY SIMILARITY.
FT	CARBOHYD	111	111		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	212	212		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	474	474		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	122	122		S -> L (IN REF. 3).
FT	CONFLICT	178	178		S -> N (IN REF. 3).
FT	CONFLICT	320	320		Y -> H (IN REF. 1).
FT	CONFLICT	474	474		N -> D (IN REF. 1).
FSQ	SEQUENCE	490 AA;	53479 MW;	07D2B03EA4D8A1A9 CRC64;	
 Query Match            24.4%; Score 454.5; DB 1; Length 490; Best Local Similarity 35.8%; Pred. No. 4.9e-32; Matches 101; Conservative 41; Mismatches 103; Indels 37; Gaps 9;					
QY	6	CAFTLLILLGTSVCG-	QPVSYSRVGGQDAAGRWPKQVSLHFDHFNYIGSGISVERLI	63	
DB	230	CSRSVRVSLRCECGSVRSQRISVGGLNAPSPGMQVSLHVQGVHVC	GSGSIITPEWI	289	
QY	64	LFAHCIOPT-----	WTFSTVTWLGSITVDGRKRVKY-----VSKIIVHPKYQDTT	112	
DB	290	VTAHCVREPLSGPRYTAF	A-----GILRQLSMFYGSRHQVEKVISHENYDSKT	339	
QY	113	A--DVALLKLSSQVTTSAIIPCLPSVTOKLATPPFCWTGWKVKESDRDYHSALQE	170		
DB	340	KNDIALMKIQTPLAFLDLVPKLPNPMPMLDLDDQCWISGWATVEKGKTS--DVLNA	397		
QY	171	AEVPIIDRAQCQLNPNTGIFLPALEPVKEKDTCAGDTQNMCKSCGSGPLSCHIDG	230		
DB	398	AMVPLEPSKCKSKYIYNLLTPAN-----ICAGPLQSGVSDCGSGGPLVTLKNG	449		
QY	231	VIQTGVSWSGECCKSL-PGVYTNVIYQKWINATISRANN	271		
DB	450	IWWLGDTSWGSCAKALRPVGYGNVYFTDWIQQM-RANS	490		
 RESULT 13 KAL_HUMAN STANDARD; PRT; 638 AA. AC P03952; DT 23-OCT-1986 (Rel. 02, Created) DT 23-OCT-1986 (Rel. 02, Last sequence update) DT 01-OCT-2000 (Rel. 40, Last annotation update) DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN) DE DE (KININOGENIN) (FLETCHER FACTOR). GN KLK3. OS Homo sapiens (Human). OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OX Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. NCBI_TaxId=9606; RN [1] RP SEQUENCE FROM N.A. RX MEDLINE=86243359; PubMed=3521732; RT Chung D.W., Fujikawa K., McMullen B.A., Davie E.W.; RT "Human plasma prekallikrein, a zymogen to a serine protease that RT contains four tandem repeats."; RL Biochemistry 25:2410-2417(1986). RN [2] RP PARTIAL SEQUENCE, AND DISULFIDE BONDS. RX MEDLINE=91152016; PubMed=1998666; RT McMullen B.A., Fujikawa K., Davie E.W.; RT "Location of the disulfide bonds in human plasma prekallikrein: the RT presence of four novel apple domains in the amino-terminal portion of RT the molecule."; RL Biochemistry 30:2050-2056(1991). CC -!- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN					

[illegible]

### Query Match

24.0%; Score 446.5; DB 1; Length 638;

Best Local Similarity 36.4%; Pred. No. 3.3e-31;  
Matches 99; Conservative 43; Mismatches 103; Indels 27; Gaps 8;

QY 2 GPAGCAFTLLLLGISVCCQPYSSRVVGGDAAAGRWPNQVSLH---FDHNFIIYGSILV 58  
DB 366 GSGSYSLRCKVGVSDCTTKI-NARIVGTTSSUGWPNQVSLQVKKLVSONHMCQGSII 424  
QY 59 SERLIITAAHCIO-----PTWTFSTVWLGSTVGDSSRRKRYVYVSKVIHPIKYO--DT 111  
DB 425 GROWILTAHCFDGIYPDPWRIYGIILNSEIT-----NKTFFSSIKELIIHQYKMSG 480  
QY 112 TADVALLKSSQVTFSSAILPCLPSVTYKQLAIPPCWVTGKGVKVESDRDYHSLQEA 171  
DB 481 SYDIALIKLQPLNYTEFOKPICLPSKADNTIYTCWVTGWTGTYTTERGET--QNTILOKA 538  
QY 172 EYPIIDRQACEQLYNPIGIFLPALEVPVIEDKICAGDTONMKDSCGDSGGPLSCHIDGV 231  
DB 539 TIPLVNECOKRYR-----DVIITKQMICAGYKEGGIDACKDGGDLVCKHSGR 589  
QY 232 WIQTGVVSWGLECG-KSLPGVTNVIYQKWI 262  
DB 590 WQLVGTISWEGGECARKEQPGVYTKVAEYIDWI 621

RESULT 15  
ID FAIL\_HUMAN STANDARD; PRT; 625 AA.  
AC P03951;  
DT 23-OCT-1986 (Rel. 02, Created)  
DT 23-OCT-1986 (Rel. 02, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE COAGULATION FACTOR XI PRECURSOR (EC 3.4.21.27) (PLASMA THROMBOPLASTIN  
DE ANTECEDENT) (PTA).  
GN Fl.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86243360; PubMed=3636155;  
RA Fujikawa K., Chung D.W., Hendrickson L.E., Davie E.W.;  
RT "Amino acid sequence of human factor XI, a blood coagulation factor  
RT with four tandem repeats that are highly homologous with plasma  
RT prekallikrein.";  
RL Biochemistry 25:2417-2424(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88107663; PubMed=2827746;  
RA Asakai R., Davie E.W., Chung D.W.;  
RT "Organization of the gene for human factor XI.";  
RL Biochemistry 26:7221-7228(1987).  
RN [3]  
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
RX MEDLINE=91152017; PubMed=1998667;  
RA McMullen B.A., Fujikawa K., Davie E.W.;  
RT "Location of the disulfide bonds in human coagulation factor XI: the  
RT presence of tandem apple domains.";  
RL Biochemistry 30:2056-2060(1991).  
RN [4]  
RP VARIANT LEU-301.  
RX MEDLINE=90046656; PubMed=2813350;  
RA Asakai R., Chung D.W., Ratnoff O.D., Davie E.W.;  
RT "Factor XI (plasma thromboplastin antecedent) deficiency in Ashkenazi  
RT Jews is a bleeding disorder that can result from three types of point  
RT mutations.";  
RN [5]  
RP Proc. Natl. Acad. Sci. U.S.A. 86:7667-7671(1989).  
RX MEDLINE=92190478; PubMed=1547342;  
RA Meijers J.C., Davie E.W., Chung D.W.;  
RT "Expression of human blood coagulation factor XI: characterization of  
RT the defect in factor XI type III deficiency.";

Blood 79:1435-1440(1992).  
-!- FUNCTION: FACTOR XI TRIGGERS THE MIDDLE PHASE OF THE INTRINSIC  
CC PATHWAY OF BLOOD COAGULATION BY ACTIVATING FACTOR IX.  
-!- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF ARG-|-ALA AND ARG-|-VAL  
CC BONDS IN FACTOR IX TO FORM FACTOR IXA.  
-!- SUBUNIT: HOMODIMER; LINKED BY A DISULFIDE BOND. AFTER ACTIVATION  
CC THE HEAVY AND LIGHT CHAINS ARE ALSO LINKED BY A DISULFIDE BOND.  
-!- PTM: ACTIVATED BY FACTOR XIIA (OR XII), WHICH CLEAVES EACH  
CC POLYPEPTIDE AFTER ARG-387 INTO THE LIGHT CHAIN, WHICH CONTAINS THE  
CC ACTIVE SITE, AND THE HEAVY CHAIN, WHICH ASSOCIATES WITH HIGH  
CC MOLECULAR WEIGHT (HMW) KININOGEN.  
-!- DISEASE: DEFECTS IN F11 ARE A CAUSE OF A BLOOD COAGULATION  
CC ABNORMALITY (ROSENTHAL SYNDROME) OCCURRING IN HIGH FREQUENCY IN  
CC ASHKENAZI JEWS.  
-!- SIMILARITY: CONTAINS 4 APPLE DOMAINS.  
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: M13142; AAA52487.1; -  
DR EMBL: M20218; AAA51985.1; -  
DR EMBL: M18296; AAA51985.1; JOINED.  
DR EMBL: M21184; AAA51985.1; JOINED.  
DR EMBL: M18298; AAA51985.1; JOINED.  
DR EMBL: M18299; AAA51985.1; JOINED.  
DR EMBL: M18300; AAA51985.1; JOINED.  
DR EMBL: M18301; AAA51985.1; JOINED.  
DR EMBL: M18302; AAA51985.1; JOINED.  
DR EMBL: M18303; AAA51985.1; JOINED.  
DR EMBL: M18304; AAA51985.1; JOINED.  
DR EMBL: M19417; AAA51985.1; JOINED.  
DR EMBL: M20217; AAA51985.1; JOINED.  
DR PIR: A27431; KFHU1.  
DR HSSP: P00763; LDPO.  
DR MEROPS: S01.213; -  
DR MIM: 134540; -  
DR MIM: 264900; -  
DR InterPro: IPR000177; -  
DR InterPro: IPR001254; -  
DR InterPro: IPR001314; -  
DR InterPro: IPR003014; -  
DR Pfam: PF00024; PAN; 4.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00005; APPLEDOMAIN.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
DR PROSITE: PS00495; APPLE; 4.  
DR Hydrolase: Serine protease; Glycoprotein; Plasma; Blood coagulation;  
KW Duplication; Signal; Disease mutation.  
FT SIGNAL 1 18  
FT CHAIN 19 387 COAGULATION FACTOR XIA HEAVY CHAIN.  
FT CHAIN 388 625 COAGULATION FACTOR XIA LIGHT CHAIN.  
FT DOMAIN 19 104 APPLE 1.  
FT DOMAIN 109 194 APPLE 2.  
FT DOMAIN 199 284 APPLE 3.  
FT DOMAIN 290 375 APPLE 4.  
FT DOMAIN 384 625 CATALYTIC.  
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 450 450 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .).  
FT ACT\_SITE 431 431  
FT ACT\_SITE 480 480  
FT ACT\_SITE 575 575

```
FT DISULFID 20 103
FT DISULFID 29 29
FT DISULFID 46 76
FT DISULFID 50 56
FT DISULFID 110 193
FT DISULFID 136 165
FT DISULFID 140 146
FT DISULFID 200 283
FT DISULFID 226 255
FT DISULFID 230 236
FT DISULFID 291 374
FT DISULFID 317 346
FT DISULFID 321 327
FT DISULFID 339 339
FT DISULFID 380 500
FT DISULFID 416 432
FT DISULFID 514 581
FT DISULFID 545 560
FT DISULFID 571 599
FT VARIANT 301 301
FT CONFLICT 226 226
SQ SEQUENCE 625 AA; 70109 MW; 147AFA94B7709E8F CRC64;

WITH A CYS RESIDUE.

INTERCHAIN.
INTERCHAIN (BETWEEN HEAVY AND LIGHT
CHAINS).

F -> L (IN F11 DEFICIENCY).
/FTid=VAR.006622.
C -> S (IN REF. 2).
```

```
Query Match 23.9%; Score 445.5; DB 1; Length 625;
Best Local Similarity 36.7%; Pred. No. 3.9e-31;
Matches 97; Conservative 47; Mismatches 97; Indels 23; Gaps 8;

QY 8 FTLLLLGISVCGQPVYSSRVGGQDAAGRPWQVSLHF---DHNFIYGGSLVSRLLIL 64
Db 369 YTLRLCKMDNECTTKI-KPRIVGGTASVRGEWPQVTLHTTPTQRHLGGSIIGNQWIL 427
QY 65 TAAHCIOPTWTFSTYVWLGSIYVDSRRKRYI-VSKIVIHPKYQ--DTTADVALLKLS 121
Db 428 TAAHCFYGVESPKILRYISGILNQSEIKEDTSFVGQEIHHQYKMAESGYDIALKLE 487
QY 122 SOVFTFSAILPICLPSTKOLATPPFCWVTWG--KVKESSDRDYHSAQEAEPIDRQ 179
Db 488 TTVNYTDSQRPICLPKSGDRNVIYTCWVTGMYRKLDRK----IQTLOKAKIPLVTNE 543
QY 180 ACEQLYNPIGIFLPALEPVIKEDKICAGDTQNMKDSCKGDSGGPLSCHIDGVWIoTGVVS 239
Db 544 ECQKRYR-----GHKITHKMICAGYREGCKDACKGDSGGPLSKKHNEVWHLVGITS 594
QY 240 WGLECG-KSLPGVYTNVIYQKWI 262
Db 595 WGECAQRERPGVYTNVVEYVDWI 618
```

Search completed: November 22, 2001, 02:45:17  
Job time: 406 sec

**This Page Blank (uspto)**



SOURCE Xenopus laevis neurula cDNA to mRNA.  
 ORGANISM Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 Xenopodinae; Xenopus.  
 REFERENCE 1 (sites)  
 \*Yamada, K.  
 TITLE The expression control of xepsin by non-axial and planar  
 posteriorizing signals in Xenopus epidermis  
 JOURNAL Unpublished (1998)  
 AUTHORS 2 (bases 1 to 2078)  
 TITLE Direct Submission  
 JOURNAL Submitted (15-OCT-1998) to the DDBJ/EMBL/GenBank databases. Kazuo  
 Yamada, Graduate School of Human Informatics, Nagoya University;  
 Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan  
 (E-mail: yamada@info.human.nagoya-u.ac.jp, Tel: +81-52-789-2572,  
 Fax: +81-52-789-2567)  
 FEATURES  
 Location/Qualifiers  
 1..2078  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 /dev\_stage="neurula"  
 35..1204  
 /gene="xepsin"  
 35..1204  
 /note="Xepsin"  
 /gene="serine protease"  
 /codon\_start=1  
 /product="epidermis specific serine protease"  
 /protein\_id="BA8494.1"  
 /db\_xref="GI:6009515"  
 /translation="MLQVLSFLVFIHQACGVPVINSRIYGGMDKRGWPMOISLS  
 YKPSICGSLTDSWMTAAHCIDSLDVSYYTLGAYQLSAPDNSVSRGVSIK  
 HPDQYEGSSDIALIEKVPFTPIPLICLSDQVFAAGTMCWYTGWNIQEGT  
 PLISPTLQKAFVALIDSSVCGTMYTESLGIPIPFSEDFEDMVCWYKEGRIDACQGD  
 SGGLVNVNVLQGLTVSWGVCAPENRFTVYQYQDMVLTNPLVIFSEEGP  
 SVAPSGPSIAPSGPVGASTTISQTEAQSIVSIEIDKVNSTIFETAMSM  
 NNTWNETFSLVSSFTISALRINETKTIDNEAQIHACSLHIALTLVLFIRFV"  
 BASE COUNT 603 a 455 c 437 g 583 t  
 ORIGIN  
 Query Match 14.5%; Score 151; DB 8; Length 2078;  
 Best Local Similarity 53.3%; Pred. No. 7.4e-36;  
 Matches 416; Conservative 1; Mismatches 346; Indels 18; Gaps 4;  
 29 tgcctctctgtggtgggtctcagtggtgtgggaacctgtatactccagcggtgttag 88  
 |||||  
 57 TACTCATTTTCATCCACCACCAAGCGTGTGTACCAAGTATCAAAATCGCAATTGTG 116  
 |||||  
 89 gtggccaggtctgctgcagggcgctgctggtgaggtgaggtgaggtgaggtgaggtg 148  
 |||||  
 117 GGGGGATGACAGTAGAGGGGGGAATGCCCTGGCAGATAAGCCTGAGCTACAAAGTG 176  
 |||||  
 149 actttatctatgaggttccctcgtcagtgaggtgaggtgaggtgaggtgaggtgaggt 208  
 |||||  
 177 ATTCTATCTCGGGGATCGCTCTCTACAGACAGCTGGGTGAGTCTGCTCTCATCTGA 236  
 |||||  
 209 tacaaccgacctggaactcttttctatactgtgtggtgaggtgaggtgaggtgaggtg 268  
 |||||  
 237 TTGATTCG---TTGGATGTTTCATATACTACTTCTGCTGTCATATCAGCTCTCTG 293  
 |||||  
 269 actcaaggaaagctgtgaagtactactgtgtcaaaatcgtcatccatcc-----ca 319  
 |||||  
 294 CCCTGACAAATCTCAGATATCCAGAGAGTGAAGAAGTATACAAAGCACCAGATTCC 353  
 |||||  
 320 agtaccagatacaacgagcagctgctgtgtgaaactgtcctctcaagtcacattca 379  
 |||||  
 354 AATACGAGAGATCCAGTGTGACATTCCTCTAATGAATGAGAGAAACCTGTCATTTTA 413  
 |||||  
 380 cttctgcaatcctgctctatttctgtccagtgctcaaaagcaggtggcaattccacct 439  
 |||||

Db 414 CCCCTTACATCTACCATCTGCCTTCCCTCTCAGGATGTTCAATTTGCACGAGGACTA 473  
 Qy 440 ttgttggtgacccggatgggaaagttaagaaagttaagaaagttaagaaagttaagaaagt 499  
 |||||  
 Db 474 TGTGCTGGTGACTGGATGGGAATATTCAGAGAAACCCCTTTAATCATCAGCCCTAAA 533  
 Qy 500 cccttcggaagcagaagtaccattattgacccgaggtgtggaacagctctacaatc 559  
 |||||  
 Db 534 CTATTCAAGAGGAGAGTTGCCATCATAGACACTGCTGTGGGACCATGATACGAAT 593  
 Qy 560 ccactcggtatcttctccagcactggagccagtcacatcaagaagaagaagatttgtgtg 619  
 |||||  
 Db 594 CCAGCTGGGCTACATCCCGATTTCAGCT---TCATCAGAGGAGACATGGTGTGTGTG 650  
 Qy 620 gtgatactcaaaacatgaagtagttgcaaggtgattcttggagggcctctgtctgttc 679  
 |||||  
 Db 651 GTTACAGAGGAGGACGATTGATGCTGCCAGGGGACTCAGGTGGACCTTTGTATGTA 710  
 Qy 680 acattgattgtgtagtaccagacagagtagtaagctgggattagaaatgtgtgtaaa 739  
 |||||  
 Db 711 ATGTCAACAATGTCTGCTCAGCTCGGTATTTGAGCTGGGGATATGGCTGCGCAGAAC 770  
 Qy 740 c---tcttcctggaggtacacacaaatgaatctactaccacaaatggatttaagcaca 796  
 |||||  
 Db 771 CCAATAGACCTGCTGTCTATACCAAGTTCAAGTATTACAGGATTTGTTGAAACCAATG 830  
 Qy 797 t 797  
 Db 831 t 831  
 RESULT 5  
 AX082979 AX082979 942 bp DNA PAT 28-FEB-2001  
 LOCUS  
 DEFINITION Sequence 3 from Patent WO0112788.  
 ACCESSION AX082979  
 VERSION AX082979.1 GI:13184900  
 KEYWORDS  
 SOURCE synthetic construct.  
 ORGANISM synthetic construct  
 REFERENCE 1 (bases 1 to 942)  
 AUTHORS Presnell, S.R. and Taft, D.W.  
 TITLE Trypsin-like polypeptide ztrypl  
 JOURNAL Patent: WO 0112788-A 3 22-FEB-2001;  
 Zymogenetics, Inc. (US)  
 FEATURES  
 Location/Qualifiers  
 1..942  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="degenerate polynucleotide sequence of mouse ztrypl  
 (SBO ID NO:2)"  
 misc\_feature 1..942  
 /note="d = A,T,C or G"  
 BASE COUNT 160 a 87 c 153 g 144 t 398 others  
 ORIGIN  
 Query Match 12.8%; Score 132.8; DB 10; Length 942;  
 Best Local Similarity 33.7%; Pred. No. 3.4e-30;  
 Matches 252; Conservative 122; Mismatches 382; Indels 12; Gaps 3;  
 Qy 44 ggaatcagtggtgtgggaacctgtatactccagcggtgtgaggtggccaggtgctg 103  
 |||||  
 Db 92 GGATHTGWSNTGTGGNCARACAAATATACNTCYAARGTNGTNAAYGNAARGCNGING 151  
 |||||  
 Qy 104 ctgagggcgctggtggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgaggtgag 163  
 |||||  
 Db 152 ARGTTGGNAATGGCCNTGGCARGTNNSNATHYNTTYTNGGNATGTAYATHTGWSNG 211  
 |||||  
 Qy 164 gttccctcgtcagtgagaggtgtgatactgacagcagcagcagcagcagcagcagcagc 223  
 |||||  
 Db 212 GWSNNTNATHCAYCAYTGGATGTHYTNACNGCNCAYTGTGTTTCARMGNSNAARA 271  
 |||||

```

clone_end:SP6
vector_side:right
47344..53015
/note="assembly_name:Contig33
clone_end:17
vector_side:right
53116..59822
/note="assembly_name:Contig32"
59923..65341
/note="assembly_name:Contig34"
65442..70898
/note="assembly_name:Contig35"
70999..76921
/note="assembly_name:Contig36"
77022..85604
/note="assembly_name:Contig37"
85705..90872
/note="assembly_name:Contig38"
90973..97315
/note="assembly_name:Contig39"
97416..105928
/note="assembly_name:Contig40"
106029..116153
/note="assembly_name:Contig41"
116254..127251
/note="assembly_name:Contig42"
127352..136839
/note="assembly_name:Contig43"
136940..142282
/note="assembly_name:Contig44"
142383..149737
/note="assembly_name:Contig45"
149838..162513
/note="assembly_name:Contig46"
162614..178181

```

BASE COUNT	52592 a	34925 c	35928 g	51602 t	3134 others
ORIGIN					
/note="assembly_name:Contig4/"					

```

Query Match      25.6%  Score 266.2;  DB 77;  Length 178181;
Best Local Similarity 98.2%  Pred. No. 9.5e-73;
Matches 268;  Conservative 1;  Mismatches 4;  Indels 0;  Gaps 0;
ORIGIN
52592 a 34925 c 35928 g 51602 t 3134 others

```

Query Match	25.6%	Score 266.2;	DB 77;	Length 178181;
Best Local Similarity	98.2%;	Pred. No. 9.5e-72;		
Matches 268;	Conservative	1;	Mismatches 4;	Indels 0;
Gaps	0;			

Matches 268; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

211 caaccgacctggactacttttcatactgtgtgtaggatcgattacagtagtgac 270  
|| |||||||  
53928 CATAGGACCTGGCACTTTTTCATATACCTGTGGTAGGATCGATTACAGTAGTGAC 53987

53928 CATAGGACCTGGACTACTTTTTCATATACTGTGTGGCTAGGATCGATTACAGTAGGTGAC 53987

[illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

Y 331 acaacgagacatcgcccttgcttgaaactgtccctcgaagtcacccttcctcgcac 390  
|||||:|||||

b 54048 ACAACGGCAGACGTCGCCCTTGTTGAAACTGTCCTCTCAAGTCACCTTCACTTCTGCCATC 54107

391 ctgcctatttgcttgccagtgtcacaaagcagttggcaattccacccttttgttgggtg 450

b 54108 CTGCCCTATTGCTTGCCCGTGCACAAAGCAGTTGGCAATTCCACCCTTTTGTGGTG 54167

451 accggaatggggaaaaatttaaggaatttcagat 483

5450  
5451  
5452  
5453  
5454  
5455  
5456  
5457  
5458  
5459  
5460  
5461  
5462  
5463  
5464  
5465  
5466  
5467  
5468  
5469  
5470  
5471  
5472  
5473  
5474  
5475  
5476  
5477  
5478  
5479  
5480  
5481  
5482  
5483  
5484  
5485  
5486  
5487  
5488  
5489  
5490  
5491  
5492  
5493  
5494  
5495  
5496  
5497  
5498  
5499

## RESULT 4

B018694	AB018694	2078 bp	mRNA	VBT	05-OCT-1999
CCUS					

DEFINITION *Xenopus laevis* Xepsin mRNA for epidermis specific serine protease,

COMPLETION DATE: 03/18/2003  
ACCESSION AB018694

ERSON AB018694.1 GI:6009514  
epidermis specific serine protease. Yersin

## RESULT 4

AB018694

LOCUS	AB018694	2078 bp	mRNA	VRT	05-OCT-1999
DEFINITION	Yersinia enterocolitica Yersinia enterocolitica mRNA for yersiniae-specific protein				

DEFINITION  
xenopus laevis xerpin mRNA for epineuritis specific serine protease,  
complete cds.

ACCESSION AB018694

VERSION AB018694.1 GI:6009514  
KEYWORDS endermis specific serine protease: Yersinia

сформировать специализированные процессы, вершин.



Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	585	31.4	389	13	Q9PVX7	Q9PVX7 xenopus lae
2	554.5	29.8	317	13	Q9DGR3	Q9DGR3 xenopus lae
3	541.5	29.1	305	11	Q9JHF7	Q9JHF7 mus musculus
4	526	28.2	314	4	Q9B6M0	Q9B6M0 homo sapien
5	523	28.1	312	4	Q9NS34	Q9NS34 homo sapien
6	522	28.0	339	11	Q9ESD1	Q9ESD1 mus musculus
7	510.5	27.4	310	11	Q9QZ29	Q9QZ29 mus musculus
8	504	27.1	342	11	Q9ER01	Q9ER01 rattus norv
9	503	27.0	342	11	Q9ES87	Q9ES87 rattus norv
10	493.5	26.5	311	11	Q9QZU7	Q9QZU7 mus musculus
11	492	26.4	297	11	Q88781	Q88781 rattus norv
12	491.5	26.4	321	4	Q9UB32	Q9UB32 homo sapien
13	490.5	26.3	306	11	Q9ER10	Q9ER10 mus musculus
14	490.5	26.3	321	4	Q9NR08	Q9NR08 homo sapien
15	488.5	26.2	317	4	Q9GZM4	Q9GZM4 homo sapien
16	477.5	25.6	321	4	Q9NR32	Q9NR32 homo sapien
17	477	25.6	300	4	Q9PZV6	Q9PZV6 homo sapien
18	472	25.3	275	4	Q9UQ11	Q9UQ11 homo sapien
19	468	25.1	273	6	Q9XSM2	Q9XSM2 ovis aries

```

clone_end:SP6
vector_side:right"
47344. .53015
/notes="assembly_name:Contig33
clone_end:T7
vector_side:right"
53116.59822
/notes="assembly_name:Contig32"
59923. .65341
/notes="assembly_name:Contig34"
65442. .70898
/notes="assembly_name:Contig35"
70999. .76921
/notes="assembly_name:Contig36"
77022. .85604
/notes="assembly_name:Contig37"
85705. .90872
/notes="assembly_name:Contig38"
90973. .97315
/notes="assembly_name:Contig39"
97416. .105928
/notes="assembly_name:Contig40"
106029. .116153
/notes="assembly_name:Contig41"
116254. .127251
/notes="assembly_name:Contig42"
127352. .136839
/notes="assembly_name:Contig43"
136940. .142282
/notes="assembly_name:Contig44"
142383. .149737
/notes="assembly_name:Contig45"
149838. .162513
/notes="assembly_name:Contig46"
162614. .178181
/notes="assembly_name:Contig47"

BASE COUNT 52592 a 34925 c 35928 g 51602 t 3134 others
ORIGIN

Query Match 25.6%; Score 266.2; DB 77; Length 178181;
Best Local Similarity 98.2%; Pred. No. 9.5e-72;
Matches 268; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 211 caacgcacgtgactcttttcatactgtgtggctaggatcgattacagtagtgac 270
Db 53928 CATAGGACCTGGACTACTTTTCATATACCTGTGTGGCTAGGATCGATTACAGTAGTGAC 53987

QY 271 tcaaggaacgctggaagtactacgtgtccaaaatcgatccatcccaagtaccagaat 330
Db 53988 TCAAGGAACGCTGGAAGTACTAGTGCTCCAAAATCGTCATCCATCCCAAGTACCAGAAT 54047

QY 331 acaagcgacagctcgctgtgtgaactgtccctcaagtcaccttcaccttcgcatc 390
Db 54048 ACAAGCGGACAGCTGCGCTGTGTGAAGTGTCCCTCAAGTCACCTTCACCTTCGCCATC 54107

QY 391 ctgacctattgtgtccagtgtcacaaagcagttggcaattccaccctttgttgggtg 450
Db 54108 CTGCGTATTGTCTTGCCAGTGTCACAAAGCAGTTGGCAATCCACCCTTTTGTGGGTG 54167

QY 451 accggatgggaaaagttaaggaagttcagat 483
Db 54168 ACCGGATGGGAAAAGTTAAGGAAAGTTTCAGGT 54200

RESULT 4
AB018694 2078 bp mRNA VRT 05-OCT-1999
LOCUS Xenopus laevis Xepsin mRNA for epidermis specific serine protease,
DEFINITION complete cds.
ACCESSION AB018694
VERSION AB018694.1 GI:6009514
KEYWORDS epidermis specific serine protease; Xepsin.

```

This Page Blank (uspic)

SOURCE Xenopus laevis neurula cDNA to mRNA.  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
Xenopodinae; Xenopus.  
REFERENCE 1 (sites)  
AUTHORS Yamada.K.  
TITLE The expression control of xepsin by non-axial and planar  
posteriorizing signals in Xenopus epidermis  
JOURNAL Unpublished (1998)  
REFERENCE 2 (bases 1 to 2078)  
AUTHORS Yamada.K., Takeshima.K. and Takabatake.T.  
TITLE Direct Submission  
JOURNAL Submitted (15-OCT-1998) to the DDBJ/EMBL/GenBank databases. Kazuto  
Yamada, Graduate School of Human Informatics, Nagoya University;  
Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan  
(E-mail:yamada@info.human.nagoya-u.ac.jp, Tel:+81-52-789-2572,  
Fax:+81-52-789-2567)

FEATURES  
Source

1..2078  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/dev\_stage="neurula"  
35..1204  
/gene="Xepsin"  
35..1204  
/gene="Xepsin"  
/note="Xepsin"  
/codon\_start=1  
/product="epidermis specific serine protease"  
/protein\_id="BAA8494.1"  
/db\_xref="GI:6009515"  
/translation="MLQYLSFVLIHHQAGVPVISNRIVGSMDSKREWPWQISL  
YKSDISGSLSDTWMTAAHCDISLVSYYVYLGAYQLSAPDNSTVSRGKSLTK  
HPDFYEGSGDIALIELEKPTFTYPIPLPSQDQVFAAGTCWVFGWNIQEGT  
PLISPTIOKAESVAIDSSVCGTVPYESSLYIPDFIOEDMVCAGYKRGIDACGD  
SGPLVNNVNLGLGIVSWGCGAEPNRPVTKVQYQDLKTNVPLVFSREGP  
SVAPSGPSIAFSPGSLGPRGVASTTISQTEAQSNSIEIDKNTTIFEFKMSM  
NNWTNFTSLVSSSTALRINEKTDIDNEAQIHACSLHTLILYLIFRFV"  
BASE COUNT 603 a 455 c 437 g 583 t  
ORIGIN

Query Match 14.5%; Score 151; DB 8; Length 2078;  
Best Local Similarity 53.3%; Pred. No. 7.4e-36;  
Matches 416; Conservative 1; Mismatches 346; Indels 18; Gaps 4;  
29 tgctctctctggggtcagtcagtggtggcgaacctgtatctactccagcgcgttgtag 88  
|||||  
57 TACTCATTTTCACCCACCACCAAGCCTGTGTGTACAGTAATCAATCGCATGTG 116  
|||||  
89 gtggcaggatgctgctgcagggcgctggccttgccaggtcgacgtcacctttgaccaca 148  
|||||  
117 GGGGATGGACATAGAGGGGGGATGGCTTGGCAGATAAGCCTGACGTACAAAAGTG 176  
|||||  
149 actttatctatggaggtccctcgtcagtgagaggttgatactgacagcagcactgca 208  
|||||  
177 ATTCTATCTCGGGGATCGCTCTTACACAGACAGCTGGGTGATGCTGCTCACTGCA 236  
|||||  
209 tacaacgacactgactcttttctatctactgtgtggtcagtgattacagtaggtg 268  
|||||  
237 TTGATTGG---TTGGATGTTTACATATACTGTTTATCTCGGTGCATATACGCTCTCTG 293  
|||||  
269 actcaaggaacggtggaagtactacgtgtccaaactcgctaccatcc-----ca 319  
|||||  
294 CCCCTGACATTTACAGATATCAGAGGAGTGAAAGATATAAACAAGCACCAGATTTC 353  
|||||  
320 agtaccagaatacaagggcagactgccttctgttgaactgtctctcaagtcacttca 379  
|||||  
354 AATACGAAGGATCCAGTGGTGACATGCTCTAATTTGAACCTGGAGAAACCTGCTACTTTA 413  
|||||  
380 ctctcgccactcctgcctattgtgtgccagtgctcaacaagcaggtggcaattccacct 439  
|||||

Db 414 CCCTTACATCCTACCCATCTGCCTTCCCTCTCAGGATGTTCAATTTGCAGCAGGACTA 473  
QY 440 ttgtgtggtgaccggtatggggaagtttaagaaagttcagatagattaccattctg 499  
|||||  
Db 474 TGTGCTGGGTGACTGTGATGGGAAATATTCAAGAAGAAACCCCTTTAATCAGGCCCTAAAA 533  
QY 500 cccttcaggaagcagaagtaccattatgacgcgagcgttctggaacagctctacaatc 559  
|||||  
Db 534 CTATTTCAGAGGAGGAGTGGCCATCATAGACAGATTCAGTCTCTGGGACCATGTACGAAT 593  
QY 560 ccatcggtatctcttctgcccagcactggagccagtcatacgaaggaagatttctgctg 619  
|||||  
Db 594 CCAGCTTGGGCTACATCCCTGATTTTCAGCT---TCATTTCAGGAGGACATGCTGTGCTG 650  
QY 620 gtgatactcaaacatgaagagatagttgcaaggttattctgagggcctctgctgctc 679  
|||||  
Db 651 GTTACAGGAGGAGGACGATTTGATGCTGCCAGGGGAGCTCAGGTGGAGCTCTTGTGATGA 710  
QY 680 acattgatggtgtatggtaccagacagcaggtagtagcgtggggttagaattggttaaat 739  
|||||  
Db 711 ATGTCAACATGCTGGCTTCAGCTCGGTATTCAGCTGGGATATGGCTGGCAGAAC 770  
QY 740 c---tcttctgaggtctacaccaaagttaattctactacacaaatgatttaagcacta 796  
|||||  
Db 771 CCAATAGACCTGCTGTATACCAAAAGTTCAAGTATTACAGGATTTGTTGAAAACCAATG 830  
QY 797 t 797  
Db 831 T 831  
RESULT 5  
AX082979 942 bp DNA PAT 28-FEB-2001  
LOCUS Sequence 3 from Patent WO0112788.  
DEFINITION AX082979  
ACCESSION AX082979  
VERSION AX082979.1 GI:13184900  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 942)  
AUTHORS Presnell,S.R. and Taft,D.W.  
TITLE Trypsin-like polypeptide ztrypl  
JOURNAL Patent: WO 0112788-A 3 22-FEB-2001;  
ZymoGenetics, Inc. (US)  
FEATURES Location/Qualifiers  
Source 1..942  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="degenerate polynucleotide sequence of mouse ztrypl  
(SEQ ID NO:2)"  
misc\_feature 1..942  
/note="n = A,T,C or G"  
BASE COUNT 160 a 87 c 153 g 144 t 398 others  
ORIGIN  
Query Match 12.8%; Score 132.8; DB 10; Length 942;  
Best Local Similarity 33.7%; Pred. No. 3.4e-30;  
Matches 252; Conservative 122; Mismatches 362; Indels 12; Gaps 3;  
QY 44 ggatctcagtggtggggaacctgtatctccagcgcgtttagtggtggcagagatgctg 103  
|||||  
Db 92 GGATHTGGWSNTGYGNCARACNAAYATHACNTGYAARGTNGTNAAYGGNAARSCNGTNG 151  
QY 104 ctgcagggcgctggccttggtgaggttcagcctacactttgaccacaaactttatctatgag 163  
|||||  
Db 152 ARGNTGGNAARTGCCNTGCCARGTNWSNATHVTNTTYTNGGNATGTAYATHITGYWSNG 211  
QY 164 gtccctcgtcagtgagaggttgaactgacagcagcacactgacacacacccactgga 223  
|||||  
Db 212 GNWSNTNATHCAICAYCAITGGATHYTNACNGCNCNCATGYITNCARMGNSNAARA 271

This Page Blank (uspto)



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☐ FADED TEXT OR DRAWING

☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☐ LINES OR MARKS ON ORIGINAL DOCUMENT

☒ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

... Page Blank (uspto)